



## Prior-Guided gated convolutional networks for rainstorm forecasting

Tong Zhang <sup>a</sup>, Jie Liu <sup>b,\*</sup>, Chulin Gao <sup>a</sup>, Peixiao Wang <sup>a,c</sup>, Liang Leng <sup>d</sup>, Yanjiao Xiao <sup>d</sup>

<sup>a</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

<sup>b</sup> School of Information Science and Engineering, University of Jinan, Jinan 250022, China

<sup>c</sup> State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

<sup>d</sup> China Meteorological Administration Basin Heavy Rainfall Key Laboratory/Hubei Key Laboratory for Heavy Rain Monitoring and Warning Research, Institute of Heavy Rain, China Meteorological Administration, Wuhan 430205, China

### ARTICLE INFO

This manuscript was handled by Yuefei Huang, Editor-in-Chief, with the assistance of Haiyun Shi, Associate Editor

#### Keywords:

Rainstorm events  
Prior-informed rainstorm forecasting  
Substantial derivative  
Spatio-temporal patterns

### ABSTRACT

Accurate rainstorm forecasting is crucial for the sustainable development of human society. Recently, machine learning-based rainstorm prediction methods have shown promising results. However, these methods often fail to adequately consider the prior knowledge of rainstorms and do not explicitly account for the dynamic spatio-temporal patterns of rainstorm events. This study introduces a novel end-to-end prior-informed rainstorm forecasting model that incorporates both fundamental physical priors and the spatio-temporal development patterns of rainstorms. The model utilizes a gated convolutional encoder-decoder network to effectively represent the spatio-temporal patterns of rainstorm events. A key component of the representation network is the *Substantial Derivative-Guided gated convolutional Unit (SDGIU)*, which updates latent states under the constraints of physical priors. Additionally, an integrated loss function is designed to minimize reconstruction errors on multiple scales and facilitate the generation of forecasts that reproduce the actual spatio-temporal patterns of rainstorm formation, development and dissipation. Experimental results on two reanalysis datasets show that the proposed forecasting model outperforms competing state-of-the-art baselines by at least 19.7% (15.0%) in overall Critical Success Index (Heidke Skill Score). Qualitative analysis indicates that the proposed model can generate predictions that are both physically consistent and spatially-temporally coherent.

### 1. Introduction

Rainstorms are typical extreme weather events that usually pose severe risks to agricultural production, urban infrastructure, and human life (Shortridge, 2019; Zhang et al., 2020). With the intensification of global climate change, rainstorms are occurring more frequently and with greater intensity (IPCC, 2022). In the future, we are likely to face more frequent and severer rainstorm events (Ornes, 2018). Therefore, accurate rainstorm forecasting plays a crucial role in providing critical risk information and decision-making suggestions for disaster prevention and mitigation, which is of great importance for the sustainable development of human society (Wu et al., 2021; Li et al., 2023). This study focuses on the prediction of rainstorms. Rainstorms are typically characterized by heavy rainfall that exceeds a certain threshold within a short period of time. However, there is no universally recognized threshold for defining rainstorms. The definitions of rainstorm are usually depend on its intensity and duration, following local standards

set by government agencies or academia.

Although advances in Numerical Weather Prediction (NWP) methods have greatly improved the prediction skills of meteorologists (Bauer et al., 2015), current NWP operational models face challenges in rainstorm forecasting due to high computational costs and errors introduced by sub-grid parameterizations (Palmer et al., 2005; Sun, 2005; Buehner and Jacques, 2020). Data-driven machine learning methods, which can capture complex spatio-temporal patterns and emulate nonlinear dynamics, have shown potential in weather modeling (Kashinath et al., 2021) and precipitation forecasting (Ravuri et al., 2021; Bi et al., 2023). Specifically, deep learning methods have been extensively used in precipitation forecasting (Shi et al., 2017; Ravuri et al., 2021; Liu et al., 2022). However, deep learning-based precipitation forecasting models may perform poorly for rainstorm forecasting due to the spatial and temporal heterogeneity of rainstorm occurrence, especially when making predictions over long time horizons (Ravuri et al., 2021). For example, convolutional networks may not be well-suited for capturing the spatial dependency of heavy rainfall (Bai

\* Corresponding author.

E-mail address: [ise\\_liuj@ujn.edu.cn](mailto:ise_liuj@ujn.edu.cn) (J. Liu).

Nomenclature		Previous studies
<i>Index of notations</i>		
$(x_i, y_i)$	Grid location	NWP Numerical Weather Prediction
$n_w, n_h$	Number of grids in $x$ or $y$ direction in the study region	TrajGRU, ConvGRU( <a href="#">Shi et al., 2017</a> ) Trajectory/Convolutional Gate Recurrent Unit
$R_t, \hat{R}_t$	Vectors of actual precipitation and predicted precipitation at time $t$ for all grids.	Simmim-Swin ( <a href="#">Xie et al., 2022</a> ) Swin Transformer based Simple framework for Masked Image Modeling
$A_t, \hat{A}_t$	Vectors of actual atmospheric attributes and reconstructed atmospheric attributes at time $t$ for all grids.	TCLR( <a href="#">Dave et al., 2022</a> ) Temporal Contrastive Learning for video Representation
$r_{xy,t}$	Precipitation at time $t$ and location $(x, y)$ .	CNN ( <a href="#">Zhou et al., 2019</a> ) Convolutional Neural Network for weather prediction
$H_{lt}$	Hidden state of the $l$ th layer at time $t$ .	FourCastNet ( <a href="#">Pathak et al., 2022</a> ) Fourier Forecasting Neural Network for weather prediction
$F_t$	Spatio-temporal statistics of the rainstorms.	STGCN ( <a href="#">Song et al., 2020</a> ) Spatial-Temporal Synchronous Graph Convolutional Networks
$S_{lt}, \hat{S}_{lt}$	Intermediate representative features and reconstructed features of the $l$ th layer at time $t$ .	
$U_l, V_l, Res_l$	Learnable gating parameters in SDGiU of the $l$ th layer (the velocity components of the hidden state in $x, y$ directions and the residual matrix)	
$Z_l, R_l$	Update gate and reset gate in ConvGRU.	
$P, P_{lt}$	Attributed spatio-temporal affinity matrix of all atmospheric events; attributed spatio-temporal affinity matrix between the current event and similar historical events.	
$p_{ij}$	Attributed spatio-temporal similarity between $event(x_i, y_i, t_i)$ and $event(x_j, y_j, t_j)$ ; an element of $P$ or $P_{lt}$ .	
$E_{lt}$	Event representations of $l$ layer at time $t$ for all grids	
$E_{lt}^{sim}, e_{l,(x_j,y_j,t_j)}^{sim}$	Representations of similar events of $l$ layer at time $t$ for all grids; elements of $E_{lt}^{sim}$ .	
$\tilde{E}_{lt}^{sim}, \tilde{e}_{l,(x_i,y_i,t_i)}^{sim}$	Integrated pattern-aware embedding of $l$ layer at time $t$ for all grids; elements of $\tilde{E}_{lt}^{sim}$ .	
$\tilde{E}_{lt}$	Activated prior-informed embedding of $l$ layer at time $t$ .	
$Gate_e, Gate_p$	Learnable activation gates for the embeddings of current events and historical events, respectively.	
$L_{rec}, L_{stc}, L_{com}$	Reconstructed loss, spatio-temporal coherence loss and the overall representation-prediction loss.	
		<i>The proposed model and its variants</i>
		PGCN Prior-guided Gated Convolutional Network
		SDGiU Substantial Derivative-Guided gated convolutional Unit
		SDGiU-EDN SDGiU based Encoder Decoder representation Network
		MSE loss Mean-Squared Error loss
		-R-SDGiU SDGiU-EDN without SDGiUs
		-SDGiU PGCN without SDGiUs
		-STCLoss PGCN without Spatial Temporal Coherence loss
		-APIE PGCN without spatio-temporal convolution and the prior activation layers (Activated Prior-Informed Embeddings are not produced and used in forecasting)
		<i>Data and performance metrics</i>
		MAE Mean Squared Error
		CSI Critical Success Index
		HSS Heidke Skill Score
		POD Probability of Detection
		MAR Miss Alarm Rate
		FAR False Alarm Rate
		ERA5 5th generation European Centre for Medium-range Weather Forecasts Reanalysis dataset
		NCEP National Centre for Environmental Prediction Climate Forecast System version 2

[et al., 2022](#)).

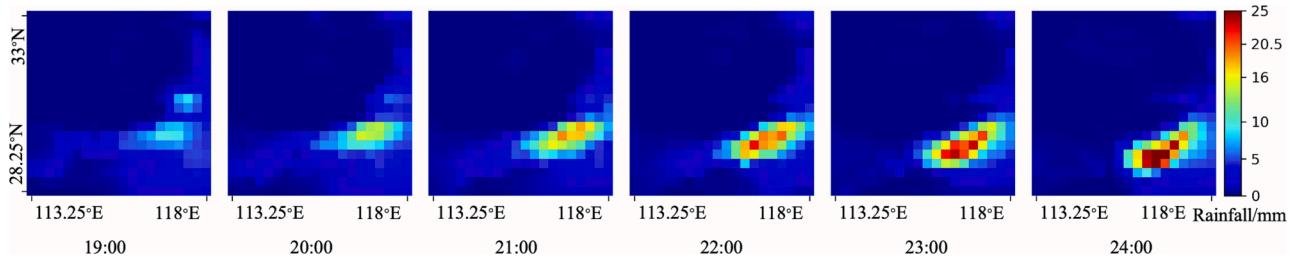
Deep learning methods have been specifically developed for rainstorm prediction. However, many of these studies either directly apply or simply extend the original deep learning models for severe convective weather or heavy rainfall prediction, such as convolutional neural networks ([Zhou et al., 2019](#)), UNet-based models ([Hess and Boers, 2022](#)), and location-refining neural network models that integrate the optical flow methods ([Huang et al., 2022](#)). A few recent deep learning-based predictors have achieved promising results by explicitly modeling the spatio-temporal correlations ([Bai et al., 2022; Zhang et al., 2022](#)) and multi-scale spatio-temporal variations of rainstorms ([Yang and Yuan, 2023](#)).

In recent years, the concept of physics-informed machine learning has been proposed to incorporate prior knowledge into machine learning models. This approach offers improved sample efficiency, generalization ability, interpretability, and convergence rates ([Karniadakis et al., 2021; von Rueden et al., 2021](#)). Significant progress has been made in developing physics-informed machine learning models for weather and climate modeling ([Kashinath et al., 2021](#)). Integrating deep learning models with equations of atmospheric motion has been shown to be beneficial for heavy rainfall nowcasting ([Ritvanen et al., 2023](#)). However, most existing data-driven forecasting methods do not adequately consider the prior knowledge of rainstorms, nor do they explicitly account for the dynamic development patterns of rainstorms.

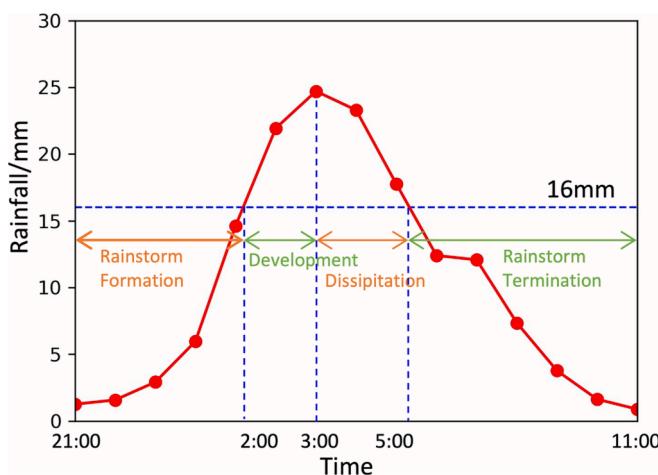
Therefore, these methods do not generalize well to untrained scenarios and may not produce accurate and physically consistent predictions based on sparse and limited rainstorm samples.

The objectives of this study are twofold: 1) to explore the possibility of integrating physical priors with deep learning models to improve the performance of rainstorm prediction; 2) to develop prior-guided deep learning-driven forecasting models that can capture the development patterns of rainstorms and generate predictions that are both physically consistent and spatially-temporally coherent. To achieve these goals, this study develops a comprehensive prior-guided deep learning model that incorporates the physical priors of rainstorms in the form of governing equations and spatio-temporal statistics. This study proposes to learn prior-guided representations for rainstorm events based on the substantial derivative equation, capturing the development patterns of rainstorms in space and time. Based on these learned event representations, an event affinity graph and a gated convolutional network are developed for rainstorm forecasting. While there are very few prior-informed machine learning-driven rainstorm forecasting models, this study explicitly integrates two different types of physical priors and develops a comprehensive prior-guided deep learning framework that effectively learns the spatio-temporal patterns of rainstorm events through a series of specially designed components.

Compared to relevant studies, the proposed model can better use the prior knowledge from historical rainstorm samples and better predict



**Fig. 1.** Spatial changes of the coverage of a rainstorm from 19:00 to 24:00 on July 5, 2018.



**Fig. 2.** Variation of precipitation with time during a rainstorm from 21:00 on July 5 to 11:00 on July 6, 2018.

the spatio-temporal development process of rainstorm events, as demonstrated by the test results on two reanalysis datasets. The proposed rainstorm forecasting model can be incorporated into operational flood warning systems to mitigate flood-related damages and provide timely weather information for travelers and residents (Piadeh et al., 2022). It can also be integrated with hydraulic simulation models to improve watershed management (Kuriqi and Ardiçlioğlu, 2018). The main contributions of this study are summarized as follows:

- 1) A *Prior-guided Gated Convolutional Network* model is proposed for rainstorm forecasting. The proposed model leverages the substantial derivatives and spatio-temporal statistics of atmospheric dynamics as physical priors. These priors are integrated and activated as meaningful inductive biases to improve the sample efficiency and generalizability of data-driven forecasting models;
- 2) A novel gated convolutional encoder-decoder network is developed to effectively represent the spatio-temporal patterns of rainstorm events. This representation network is built upon a specially designed *Substantial Derivative-Guided gated convolutional Unit* (SDGiU), which plays a critical role in integrating physical priors with the gated convolutional network;
- 3) An integrated loss function is designed to minimize the reconstruction errors on multiple scales, the prediction error loss, and the spatio-temporal coherence loss, facilitating the generation of forecasting results that reproduce the actual spatio-temporal patterns of rainstorm formation, development and dissipation.

## 2. Definitions and problem formulation

### 2.1. Rainstorm event and process

There is no globally recognized standard for rainstorms. Since this

study focuses on rainstorms in China, this study follows the official standard set by the China Meteorological Administration, which states that “rainfall of 16 mm or more per hour is considered a rainstorm” (China Meteorological Administration, 2012). In this study, a heavy rainstorm is defined as hourly precipitation exceeding 20.5 mm. Formally, rainstorm events are defined as follows:

**Definition 1. Rainstorm event.** A “rainstorm event”  $event_i(x_i, y_i, t_i)$  refers to a phenomenon in which heavy rainfall occurs at a specific time  $t_i$  and location  $(x_i, y_i)$ . The study area is divided into  $n_w \times n_h$  grids and the precipitation over a specific time period (i.e., one hour) before time  $t$  for all grids is denoted as  $R_t \in \mathbb{R}^{n_w \times n_h}$ . The precipitation over a one-hour time period before time  $t$  for a grid  $(x, y)$  centered at longitude “lon.” and latitude “lat.” is denoted as  $r_{x,y,t}$ . When  $r_{x,y,t} < 16$  mm, it is considered that no rainstorm event has occurred on the grid  $(x, y)$  at time  $t$ . When  $r_{x,y,t} \geq 16$  mm, it is considered that a rainstorm event has occurred on the grid  $(x, y)$  at time  $t$ . When  $r_{x,y,t} \geq 20.5$  mm, it is considered that a heavy rainstorm event has occurred on the grid  $(x, y)$  at time  $t$ .

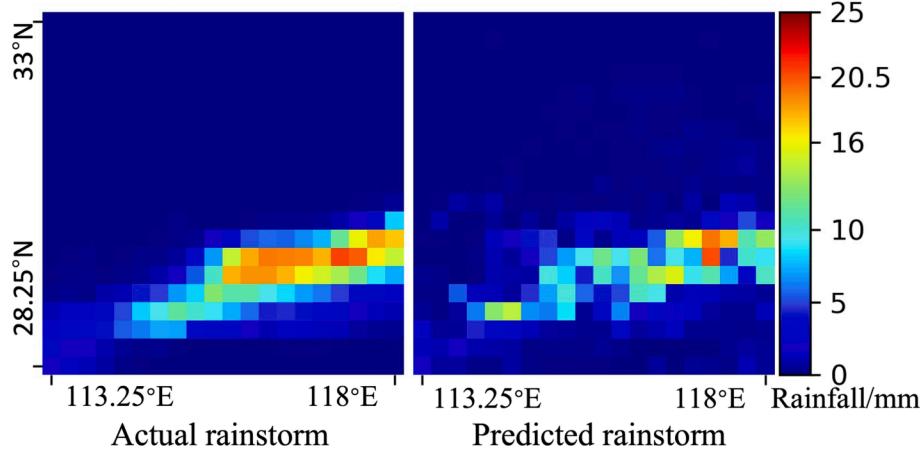
$$\begin{cases} r_{x,y,t} < 16\text{mm}, & \text{no rainstorm} \\ 16\text{mm} \leq r_{x,y,t} < 20.5\text{mm}, & \text{rainstorm events} \\ 20.5\text{mm} \leq r_{x,y,t}, & \text{heavy rainstorm events} \end{cases} \quad (1)$$

**Definition 2. Rainstorm process.** The rainstorm process refers to the spatial and temporal process of rainstorm formation, development, dissipation and termination in a specific region. It is a collection of rainstorm events that describe spatio-temporal changes in rainstorm coverage and precipitation. Figs. 1 and 2 depict the entire process of a rainstorm in terms of spatial coverage and temporal variation of precipitation, respectively.

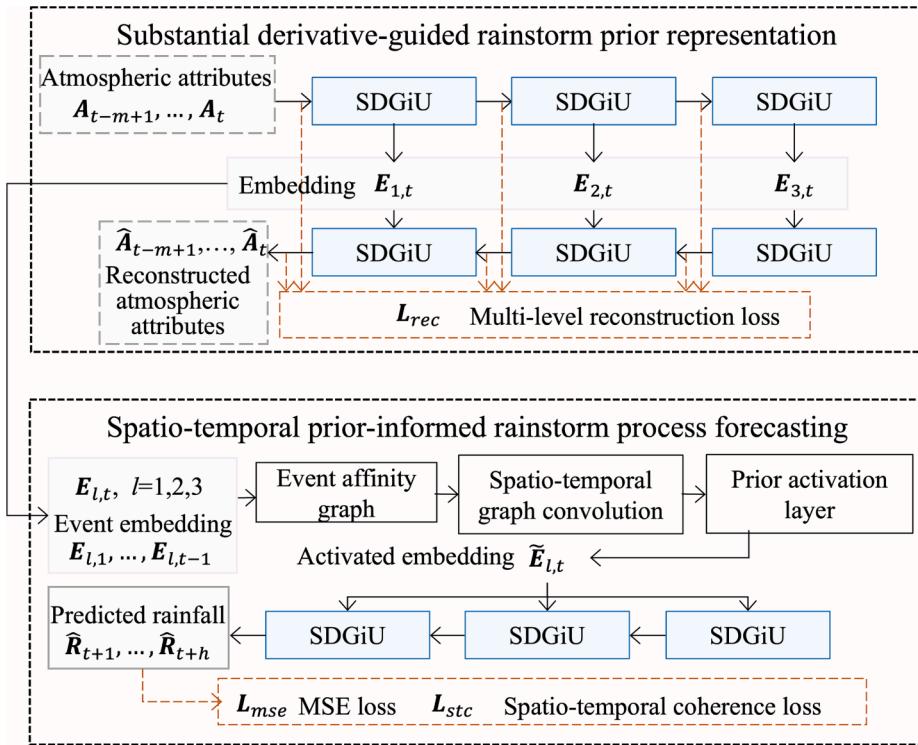
The chronologically-arranged sub-figures in Fig. 1 show the coverage of the rainstorm during the stages of formation, development, and dissipation. The formation stage occurs when rainstorm events begin to occur in the region as precipitation exceeds the threshold of 16 mm. The development stage is characterized by an increase in the number of rainstorm events in the region, as indicated by a gradual increase in the number of grid cells in which rainstorms occur. The dissipation stage begins when the number of rainstorm events decreases, eventually leading to the termination of all rainstorm events.

Fig. 2 demonstrates that when the precipitation reaches 16 mm, the location is considered to be in the rainstorm formation stage; as the precipitation continues to rise, the rainstorm enters the development stage; once the precipitation starts to fall, the rainstorm enters the dissipation stage. When the precipitation continues to decrease to less than 16 mm, the storm is considered to have ended. To generalize a reliable model, it is necessary to utilize both rainstorm events and non-rainstorm events for training the prediction model, as relying solely on a small number of rainstorm samples may not be sufficient.

The rainstorm process contains rich spatio-temporal information. As shown in Fig. 3, there are clear spatio-temporal correlations in rainfall between neighboring grid cells, with stronger correlations observed in



**Fig. 3.** Rainfall forecasts that are not consistent with typical development patterns of rainstorm. Abrupt transitions between rainstorm cells and non-rainstorm cells are observed.



**Fig. 4.** Workflow of the prior-informed gated convolutional network for short-term rainstorm process forecasting.

heavy rainstorms. Neglecting these correlations in rainstorm events may result in poor predictions that are not spatially-temporally coherent. Therefore, it is crucial to explicitly model and represent the entire rainstorm process in order to capture these rainfall correlations in space and time. While purely data-driven prediction models may fail to generate robust forecasting results that are physically consistent with real-world rainstorm development patterns, this study proposes to develop a prior-guided neural network to effectively represent historical rainstorm and non-rainstorm events, thereby improving the ability of data-driven models to capture complex spatio-temporal correlations in rainstorms.

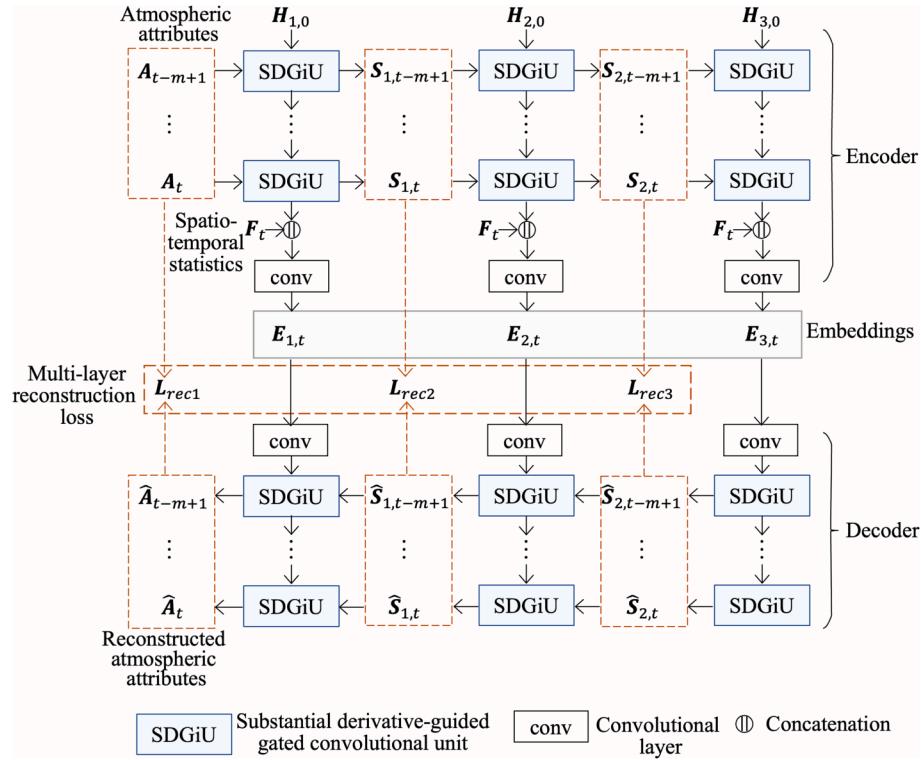
## 2.2. Rainstorm prediction problem

According to the definitions of rainstorm events and processes, the

problem of rainstorm prediction can be formally defined as follows:

$$\hat{\mathbf{R}}_{t+1}, \dots, \hat{\mathbf{R}}_{t+h} = \mathcal{F}(\mathbf{R}_{t-m+1}, \dots, \mathbf{R}_t; \mathbf{A}_{t-m+1}, \dots, \mathbf{A}_t) \quad (2)$$

where  $\hat{\mathbf{R}}_{t+1}, \dots, \hat{\mathbf{R}}_{t+h} \in \mathbb{R}^N$  denote the amounts of predicted precipitation at each of the next  $h$  time steps for all  $N$  grid cells in the study region.  $\mathbf{R}_{t-m+1}, \dots, \mathbf{R}_t$  denote the precipitations of these grid cells at the previous  $m$  time steps.  $\mathbf{A}_{t-m+1}, \dots, \mathbf{A}_t \in \mathbb{R}^{k \times N}$  are  $k$ -dimensional tensors that encode historical atmospheric attributes for all grid cells, including critical meteorological variables for rainstorm forecasting, such as temperature, precipitation, wind, and humidity.  $\mathcal{F}(\cdot)$  is a data-driven predictor that learns the spatio-temporal dynamics of rainstorm development. According to Eq.(2), grid cells are the basic spatial units for rainstorm forecasting. The output of the model is the forecasted rainfall for each grid cell for each of the next  $h$  time steps. This study sets  $m = 6$ ,  $h = 3$ ,



**Fig. 5.** Substantial derivative-guided encoder-decoder network (SDGiU-EDN).

meaning that the prediction model makes rainstorm forecasts in the next 3 hours based on meteorological data of the previous 6 hours. By applying the pre-defined rainstorm thresholds, the location, coverage, and intensity of future rainstorms can be identified, which provide essential information for studying the development of rainstorm events.

### 3. Method

#### 3.1. Overview

The substantial derivative equation is a fundamental equation for describing atmospheric flows. Since many meteorological quantities vary in time and space, the substantial derivative equations are used to represent the fundamental physical prior, i.e., how atmospheric dynamics change along the path of a rainstorm. It is assumed that the development of rainstorms follows the substantial derivative equation. The spatio-temporal statistics of rainstorms are assumed to encode spatio-temporal development patterns, and should also be used as priors and integrated into the forecasting model. The substantial derivative equation and spatio-temporal statistics of rainstorms are incorporated into a data-driven deep rainstorm forecasting model to better capture the spatio-temporal variability of meteorological quantities. The integration of substantive derivatives also contributes to the generation of physically consistent forecasts by constraining the model to obey basic physical principles that govern atmospheric dynamics.

As shown in Fig. 4, the proposed rainstorm forecasting method consists of two modules: 1) substantial derivative-guided rainstorm prior representation; and 2) spatio-temporal prior-informed rainstorm process forecasting. The underlying rationale is simple: this method first extracts meaningful representations from historical rainstorm events as physical priors, and then integrates these priors into a deep learning-based model to perform prior-informed rainstorm forecasting. The two modules are briefly described as follows:

1) Substantial derivative-guided rainstorm prior representation. This module is essentially a prior representation network that produces comprehensive representations (embeddings) of atmospheric properties and rainstorm statistics for the previous  $m$  time steps. The prior representation module is a 3-layer encoder-decoder network that includes multiple *Substantial Derivative Guided Gated Convolutional Units* (SDGiUs). These SDGiUs update the latent states in a physically-consistent manner. A multi-level reconstruction loss is proposed to retain as much spatio-temporal and attributed information as possible during representation learning.

2) Spatio-temporal prior-informed rainstorm process forecasting. Rainstorm event embeddings are used to construct an event affinity graph based on their similarity. The historical pattern-aware prior embeddings are obtained by performing spatio-temporal graph convolution on the event affinity graph, which essentially aggregates the embeddings of similar events. A prior activation layer is developed to fuse historical pattern-aware embedding with the current event embedding. The resulting activated embeddings are fed into multiple SDGiUs to forecast rainfall values for each grid cell over the next  $h$  time steps. In addition to the traditional Mean Squared Error (MSE) loss, a spatio-temporal coherence loss is proposed to enforce pattern constraints so that the predicted rainstorms follow the development patterns of typical rainstorm process.

During training, a comprehensive loss function consisting of the multi-level reconstruction loss, the MSE loss, and the spatio-temporal coherence loss is used to integrate the two modules into a unified end-to-end framework.

The entire training, validation, and testing workflow described in Fig. A1 of the Appendices. The training process of the proposed Prior-guided Gated Convolutional Network (PGCN) model mainly consists of three components, i.e., substantial derivative-guide rainstorm prior representation, spatio-temporal prior-informed rainstorm forecasting, and the overall representation-prediction loss computation, which will be elaborated in Sections 3.3, 3.4, and 3.5, respectively. The

**Table 1**

Spatio-temporal statistics of rainstorm events.

	Temporal	Spatial
Precipitation	Maximum, minimum, median, mode, mean, standard deviation, skewness, kurtosis, coefficient variation	
Rainstorm	Average increase of rainfall	Average increase of rainfalls in north-south and east-west directions
	Average rainfall amount	Average rainfall amount
	Duration of continuous rainstorm	Length and width of minimum bounding rectangle
	Total accumulated time span	Total coverage area

computation procedure of SDGiU-EDN is described in Algorithms A1 and A2 in the Appendices.

### 3.2. Substantial derivative-guided rainstorm prior representation

#### 3.2.1. Prior representation network

We develop a substantial derivative-guided prior representation network to characterize rainstorm events in a manner consistent with the physical laws of atmospheric motion, with the goal of accounting for the complex spatio-temporal nonlinear relationships of multiple atmospheric properties. The prior representation network extracts spatio-temporal dynamic features of both atmospheric and rainstorm properties, enabling reliable and generalizable rainstorm prediction. The prior representation model is an encoder-decoder network consisting of multiple SDGiUs in three layers. Therefore, the substantial derivative-guided representation network can be termed as SDGiU-EDN (i.e. Encoder-Decoder Network based on SDGiUs). The encoder takes atmospheric features as inputs and generates event representation vectors at three levels. These representation vectors contain rich spatio-temporal and attributed features for both rainstorm and non-rainstorm events. The decoder then uses these representation vectors to reconstruct atmospheric features.

As shown in Fig. 5, SDGiU-EDN has a typical encoder-decoder architecture. The encoder (decoder) consists of multiple layers, with each layer containing  $m$  recurrent SDGiUs. Each SDGiU corresponds to the original (reconstructed) atmospheric attributes at one of the previous  $m$  time steps. In the encoder, the input is the atmospheric attributes  $A_{t-m+1}, \dots, A_t \in \mathbb{R}^{c_a \times N}$  of the last  $m$  time steps. The hidden state matrix  $H_{1,0} \in \mathbb{R}^{c_h \times N}$  of layer 1 is initialized as an all-zero matrix. The outputs of the  $m$  SDGiUs in layer 1 are the embeddings of  $A_{t-m+1}, \dots, A_t$ , i.e., the spatio-temporal features of layer 1 ( $S_{1,t-m+1}, \dots, S_{1,t}$ ). The hidden states of layer 1 (i.e.,  $H_{1,t}$ ) and the spatio-temporal statistics of the rainstorms  $F_t \in \mathbb{R}^{c_f \times N}$  (see section 3.2.2 for detailed description) are concatenated and fused by convolution to obtain the event representation of layer 1 ( $E_{1,t} \in \mathbb{R}^{c_1 \times N}$ ).  $E_{1,t}$  can be used as part of the prior representation of the spatio-temporal patterns of rainstorms and as the initial hidden state of

the decoder of layer 3.

The second and third layers of the encoder follow a similar procedure as layer 1 to produce their respective event representations  $E_{2,t} \in \mathbb{R}^{c_2 \times N}$  and  $E_{3,t} \in \mathbb{R}^{c_3 \times N}$ . The event representation embeddings  $E_{1,t}$ ,  $E_{2,t}$  and  $E_{3,t}$  contain rich prior information for studying the development and evolution of rainstorms in the study region. Each layer of the decoder uses the event representation of the corresponding layer of the encoder as the initial hidden state, and reconstructs embeddings of that layer based on the reconstruction results of the previous layer. Finally the decoder reconstructs the vectors of atmospheric attributes  $\hat{A}_{t-m+1}, \dots, \hat{A}_t \in \mathbb{R}^{c_a \times N}$ .

#### 3.2.2. Incorporating additional spatio-temporal statistics of rainstorms as priors

In addition to atmospheric attributes, the spatio-temporal statistics of rainstorms are also incorporated as priors and feed into the substantial derivative-guided representation module. Table 1 shows the statistical variables used to describe the development patterns of rainstorm events. These statistics are recorded as multi-dimensional time series data for each grid cell and can be represented as  $F_t \in \mathbb{R}^{c_f \times N}$ , where  $c_f = 193$  is the dimensionality of all the statistics and  $N$  is the number of grid cells in the study region. For each grid cell, we collect temporal statistics for the previous 6 time steps, as well as spatial statistics for its  $3 \times 3$  and  $4 \times 4$  neighboring cells. Including these spatio-temporal statistics enriches the representation of rainstorm patterns and improves the efficiency of limited rainstorm samples. Additionally, the spatial morphology of rainstorm events is described using the minimum bounding rectangle, as shown in Fig. 6. Intuitively, the rainfall intensity is positively proportional to the size of the convective cloud system, which can be roughly represented by the minimum bound rectangle of a rainstorm.

#### 3.2.3. Substantial derivative-guided gated convolutional units

The proposed SDGiU extends the convolutional gated recurrent unit (ConvGRU, Shi et al., 2017) by introducing a new gating mechanism, which regulates the updating of temporal hidden states in recurrent networks. The new gating mechanism is guided by the substantial derivative equation (see Appendices 1), which enforces the physical constraints on the latent space and improve the representative capability. Eq. (A1) can be rewritten as,

$$\frac{dH}{dt} - w \frac{\partial H}{\partial z} = u \frac{\partial H}{\partial x} + v \frac{\partial H}{\partial y} + \frac{\partial H}{\partial t} \quad (3)$$

Since the total substantial derivative  $\frac{dH}{dt}$  and the convective derivative  $\frac{\partial H}{\partial z}$  of the meteorological variables in  $z$  direction are difficult to describe and compute, we combine them into the adaptive residual term  $Res$ . The computation of local derivative  $\frac{\partial H}{\partial t}$  can be further simplified by representing them as the difference between the physical quantities at two connected time steps  $H_l - H_{l-1}$ .  $\frac{\partial H}{\partial x}$  and  $\frac{\partial H}{\partial y}$  can be computed as the

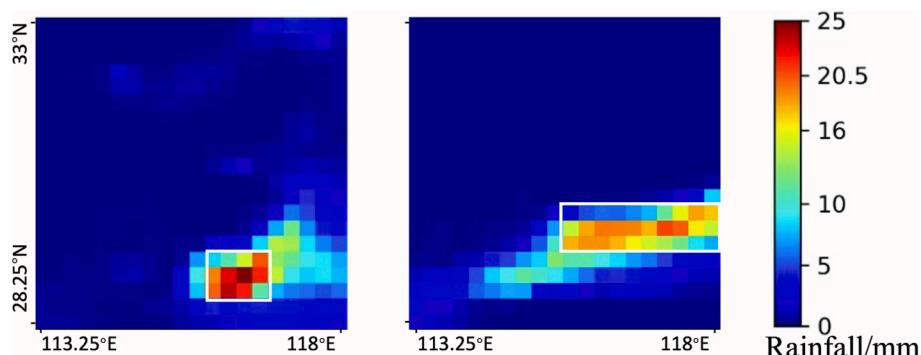
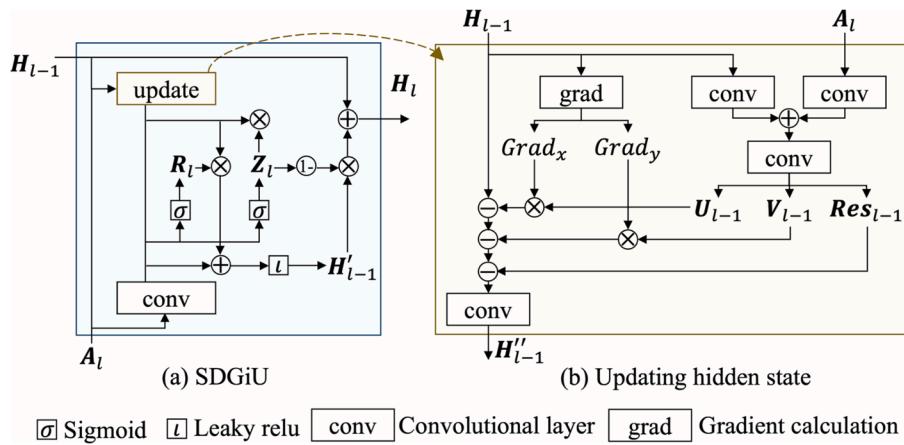
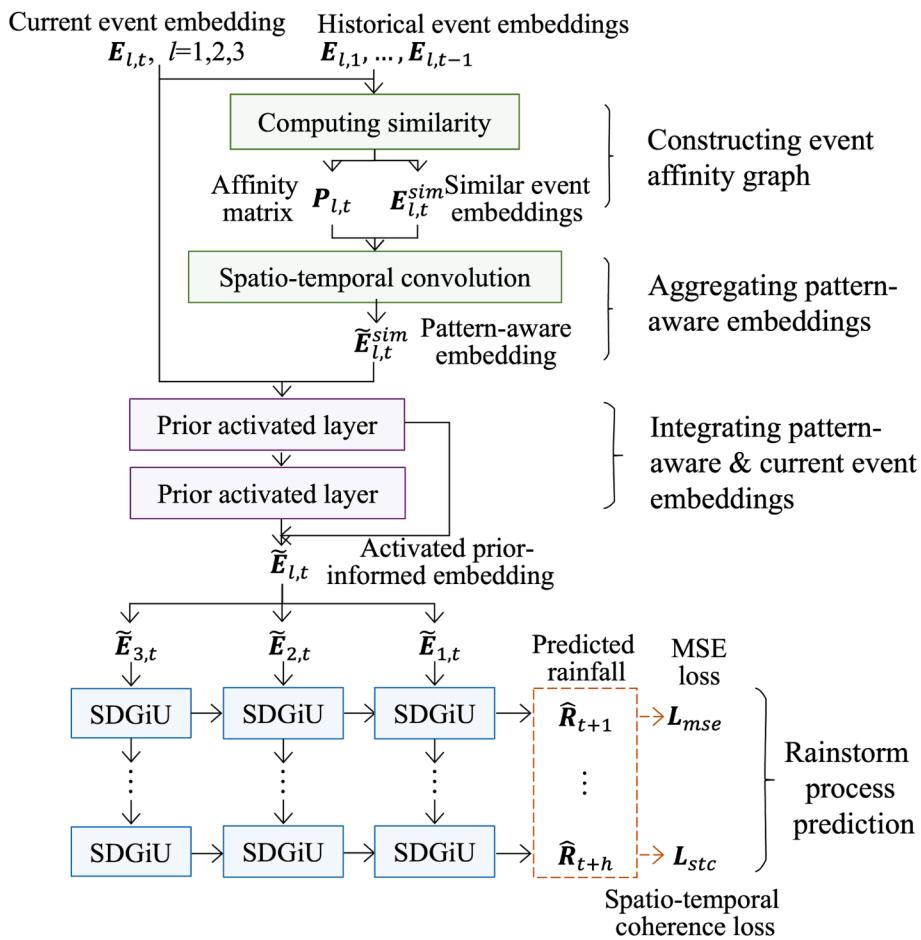


Fig. 6. Minimum bounding rectangles of rainstorms. White rectangles indicate the current spatial coverage of rainstorm events.



**Fig. 7.** Workflow of SDGiU.



**Fig. 8.** The workflow of the forecasting module

difference between the physical quantities of a grid point and the adjacent grid point in  $x$  and  $y$  directions, respectively. Practically, the computation of  $\frac{\partial H}{\partial x}$  and  $\frac{\partial H}{\partial y}$  can be done by setting the weights of convolution, i.e.,  $\text{grad}_x(H_{l-1}) = \text{conv}(H_{l-1}, \text{weight} = [1, -1])$ ,  $\text{grad}_y(H_{l-1}) = \text{conv}(H_{l-1}, \text{weight} = [1, -1]^T)$ , where  $\text{conv}(\cdot)$  denotes the convolution operation. By replacing the corresponding terms with these simplified counterparts, Eq. (A1) can be re-written as,

$$\mathbf{H}_l - \mathbf{H}_{l-1} + u \cdot \text{grad}_x(\mathbf{H}_{l-1}) + v \cdot \text{grad}_y(\mathbf{H}_{l-1}) + \mathbf{Res} = 0 \quad (4)$$

According to Eq. (4), the computation of the updated hidden states  $H'_{t-1}$  can be written as:

$$\begin{cases} U_{l-1}, V_{l-1}, Res_{l-1} = uvr(A_l, H_{l-1}) \\ H_{l-1}'' = H_{l-1} - U_{l-1}^\circ grad_x(H_{l-1}) - V_{l-1}^\circ grad_y(H_{l-1}) - Res_{l-1} \end{cases} \quad (5)$$

where  $\text{nnr}(\cdot)$  is a two-layer convolutional network,  $U_{l-1}$ ,  $V_{l-1}$ ,  $\mathbf{Res}_{l-1}$  are learnable gating parameters and  $\circ$  is Hadamard product. By integrating the simplified substantial derivative equation, this method can perform physically-constrained updating of the hidden state for the current time

step  $l$ . The results of  $\text{conv}(\bullet)$  include the residual matrix  $\mathbf{Res}$  and the velocity components of the hidden state in  $x$  and  $y$  directions ( $U, V$ ).

The embedding of the updated hidden state  $\tilde{\mathbf{H}}_{l-1}$  is fed into the ConvGRU to extract the hidden state representation  $\mathbf{H}_l$ , as described in Eq.(6),

$$\begin{cases} \mathbf{Z}_l = \sigma(\text{conv}_{az}(\mathbf{A}_l) + \text{conv}_{hz}(\tilde{\mathbf{H}}_{l-1})) \\ \mathbf{R}_l = \sigma(\text{conv}_{ar}(\mathbf{A}_l) + \text{conv}_{hr}(\tilde{\mathbf{H}}_{l-1})) \\ \mathbf{H}'_l = \iota(\text{conv}_{ah}(\mathbf{A}_l) + \mathbf{R}_l^\circ \text{conv}_{hh}(\tilde{\mathbf{H}}_{l-1})) \\ \mathbf{H}_l = (1 - \mathbf{Z}_l)^\circ \mathbf{H}'_l + \mathbf{Z}_l^\circ \mathbf{H}_{l-1} \end{cases} \quad (6)$$

$\mathbf{A}_l$  is a tensor that describes the atmospheric properties at time  $l$ .  $\mathbf{Z}_{l+1}$  denotes a update gate, which controls the retention of previous memory.  $\mathbf{R}_l$  is a reset gate that determines how the updated state  $\tilde{\mathbf{H}}_{l-1}$  is integrated with  $\mathbf{A}_l$ .  $\mathbf{H}'_l$  is the candidate hidden state. The subscripts of  $\text{conv}(\bullet)$ ,  $a, z, h, r$ , indicate the inputs and outputs of convolutional operations. For example,  $\text{conv}_{az}$  denotes the input of  $\text{conv}(\bullet)$  is  $\mathbf{A}_l$  and the output is part of  $\mathbf{Z}_l$ .  $\sigma$  is the sigmoid function and  $\iota$  is the leakyRelu function.

Fig. 7 shows the addition of a new state updating gate before performing memory updating using the classical ConvGRU (Shi et al., 2017). Fig. 7(b) depicts the detailed internal workflow of the new substantial derivative-guided updating gate.

In summary, the  $l$ th SDGiU takes the atmospheric attribute tensor  $\mathbf{A}_l$  and the hidden state  $\mathbf{H}_{l-1}$  of the previous SDGiU as input, performing substantial derivative-guided updating based on Eq. (5). Then using Eq. (6), we can derive the hidden state  $\mathbf{H}_l$  of the  $l$ th SDGiU and use it as the input for the next SDGiU. Each layer of SDGiU-EDN consists of  $m$  recurrent SDGiUs that embed the atmospheric attribute tensor  $\mathbf{A}_{t-m+1}, \dots, \mathbf{A}_t$  and an all-zero matrix  $\mathbf{H}_0$  into the hidden state  $\mathbf{H}_t$ .  $\mathbf{H}_t$  is then fused with the spatio-temporal statistics tensor  $\mathbf{F}_t$  to obtain event representation  $\mathbf{E}_t$ .

### 3.3. Spatio-temporal prior-informed rainstorm process forecasting

#### 3.3.1. The forecasting network

Based on the event embeddings results, prior-informed rainstorm process forecasting can be performed. Fig. 8 illustrates the workflow of the forecasting module. The historical event representations are integrated with the current event representations to predict the rainfall values for the next  $h$  time steps using a substantial derivative-guided gated convolutional network.

In the forecasting network, the embeddings of historical events are integrated as spatio-temporal pattern priors. As shown in Fig. 8, there are three key steps in generating activated prior-informed embeddings, which integrate the representations of both historical events and current events: (1) constructing an event affinity graph based on the similarity between the current event embeddings and all the historical events embeddings in the training dataset; 2) aggregating pattern-aware embeddings via spatio-temporal graph convolution; and 3) integrating pattern-aware embeddings with the current event embeddings  $\tilde{\mathbf{E}}_{l,t}$  ( $l = 1, 2, 3$ ) via prior activation layers. Details of these three steps are given in sections 3.3.2, 3.3.3, and 3.3.4, respectively.

The forecasting network has the same structure as the decoder of SDGiU-EDN and consists of three layers of SDGiU modules. The only difference is that each module of the forecasting network uses  $h$  SDGiU while SDGiU-EDN uses  $m$  SDGiU because the outputs of the two networks have different time steps. The output of the forecasting network is the rainfall value after  $h$  time steps in the future.

The forecasting network takes  $\tilde{\mathbf{E}}_{l,t}$  ( $l = 1, 2, 3$ ) as the only input.  $\tilde{\mathbf{E}}_{3,t}$  is fed into the first layer as the initial hidden state, which is decoded as  $\hat{\mathbf{R}}_{1,t+1}, \dots, \hat{\mathbf{R}}_{1,t+h}$  as the output of the first layer. Then the second layer uses  $\hat{\mathbf{R}}_{1,t+1}, \dots, \hat{\mathbf{R}}_{1,t+h}$  as input and produces  $\hat{\mathbf{R}}_{2,t+1}, \dots, \hat{\mathbf{R}}_{2,t+h}$ . Finally,  $\hat{\mathbf{R}}_{2,t+1}, \dots, \hat{\mathbf{R}}_{2,t+h}$  are fed into the third layer to generate predicted rainfall

values  $\hat{\mathbf{R}}_{t+1}, \dots, \hat{\mathbf{R}}_{t+h}$ .

In addition to the commonly used MSE loss, a spatio-temporal coherence loss is also developed to promote spatio-temporal consistency between the predicted results ( $\hat{\mathbf{R}}_{t+1}, \dots, \hat{\mathbf{R}}_{t+h}$ ) and the actual rainfalls ( $\mathbf{R}_{t+1}, \dots, \mathbf{R}_{t+h}$ ). Details of the loss function is introduced in section 3.4.

#### 3.3.2. Constructing event affinity graph

To address the challenge of sparse and limited rainstorm samples, an event affinity graph is constructed based on the similarity between the embeddings of each event and its historical events.

**Definition 3. Attributed spatio-temporal similarity.** For any two atmospheric events  $\text{event}_i(x_i, y_i, t_i)$  and  $\text{event}_j(x_j, y_j, t_j)$ , we define a comprehensive attributed spatio-temporal similarity metric  $p_{ij}$ , which captures similarities in the dimensions of attribute, space, and time.

$$p_{ij} = \text{cosine}\left(\mathbf{e}_{(x_i, y_i, t_i)}, \mathbf{e}_{(x_j, y_j, t_j)}\right) \quad (7)$$

where  $\mathbf{e}_{(x_i, y_i, t_i)}$  is the learned event embedding of SDGiU-EDN. Inter-event similarity can be directly computed based on the learned event embeddings. Note that  $\text{Event}_i(x_i, y_i, t_i)$  and  $\text{Event}_j(x_j, y_j, t_j)$  are not necessarily adjacent in space or time.

**Definition 4. Event affinity graph.** An event affinity graph  $\mathcal{G}$  is defined on the  $n_w \times n_h$  grid of the study region, spanning the time period  $[0, t]$ . Each node in the graph represents an atmospheric event (including both rainstorm and non-rainstorm events). Links between nodes represent virtual pairwise spatio-temporal correlations between events. Therefore,  $\mathcal{G}$  has a total of  $N = |t| \times n_w \times n_h$  spatio-temporal nodes. Each spatio-temporal node  $v_i$  denotes an atmospheric event  $\text{Event}(x_i, y_i, t_i)$ , which occurs at time  $t_i$  in the grid cell  $(x_i, y_i)$ . The attribute of each node is its embedding  $\mathbf{e}_{(x_i, y_i, t_i)}$ , which is computed by SDGiU-EDN. The attributed spatio-temporal affinity matrix  $\mathbf{P} \in \mathbb{R}^{N \times N}$  measures the comprehensive correlations between events. Each element  $p_{ij}$  describes the attributed spatio-temporal similarity between  $\text{event}(x_i, y_i, t_i)$  and  $\text{event}(x_j, y_j, t_j)$ . Note that in order to produce a compact graph, each node keeps only the most  $n_{ref}$  similar events as neighbors. In this way, the attributed spatio-temporal affinity matrix becomes  $\mathbf{P}_{l,t} \in \mathbb{R}^{(n_w \times n_h) \times n_{ref}}$ .

#### 3.3.3. Deriving pattern-aware embedding via spatio-temporal graph convolution

Based on the constructed event affinity graph, this study uses spatio-temporal graph convolution to aggregate meaningful embeddings of similar events into integrated pattern-aware embeddings. These embeddings incorporate the development pattern priors of rainstorms and can be used to support accurate rainstorm prediction. Each node in the event affinity graph  $\mathcal{G}$ , has  $n_{ref}$  neighbors that have complex attributed spatio-temporal correlations with itself. To improve the representativeness of nodal embeddings, we can use spatio-temporal convolution to aggregate the embeddings of neighboring nodes.

**Definition 5. Spatio-temporal convolution.** Based on the event affinity graph  $\mathcal{G}$  and its associated attributed spatio-temporal affinity matrix  $\mathbf{P}_{l,t} \in \mathbb{R}^{(n_w \times n_h) \times n_{ref}}$ , we can compute the integrated pattern-aware embeddings at the node level by the following equation:

$$\tilde{\mathbf{E}}_{l,t}^{\text{sim}} = \mathbf{P} \mathbf{E}_{l,t}^{\text{sim}} \mathbf{W}_g + \mathbf{b}_g \quad (8)$$

where  $\mathbf{W}_g \in \mathbb{R}^{c_l \times c_l}$ ,  $\mathbf{b}_g \in \mathbb{R}^{c_l}$  denote the learnable weights.  $\mathbf{E}_{l,t}^{\text{sim}} \in \mathbb{R}^{n_{ref} \times c_l}$  represents the embeddings of neighboring nodes. Eq. (8) can be written in vector matrix form:

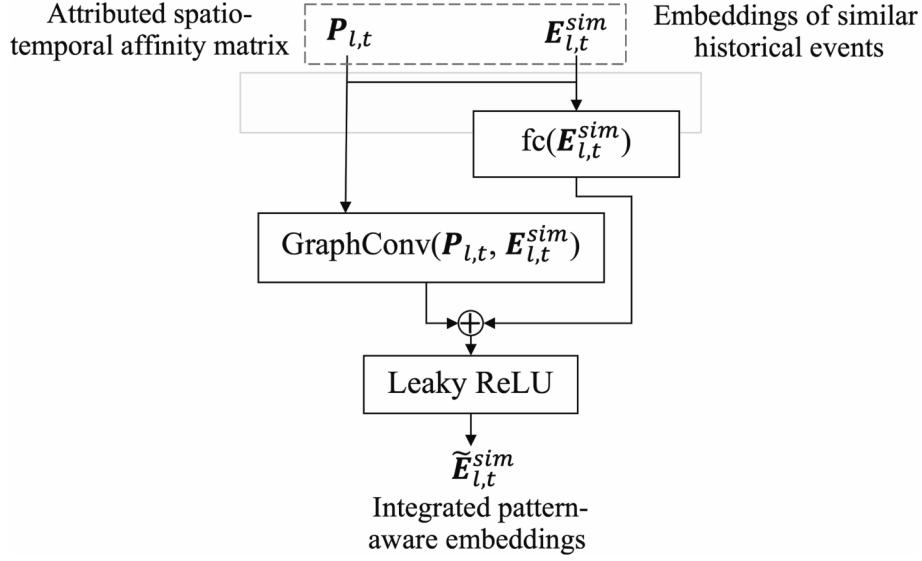


Fig. 9. The computation of integrated pattern-aware embeddings.

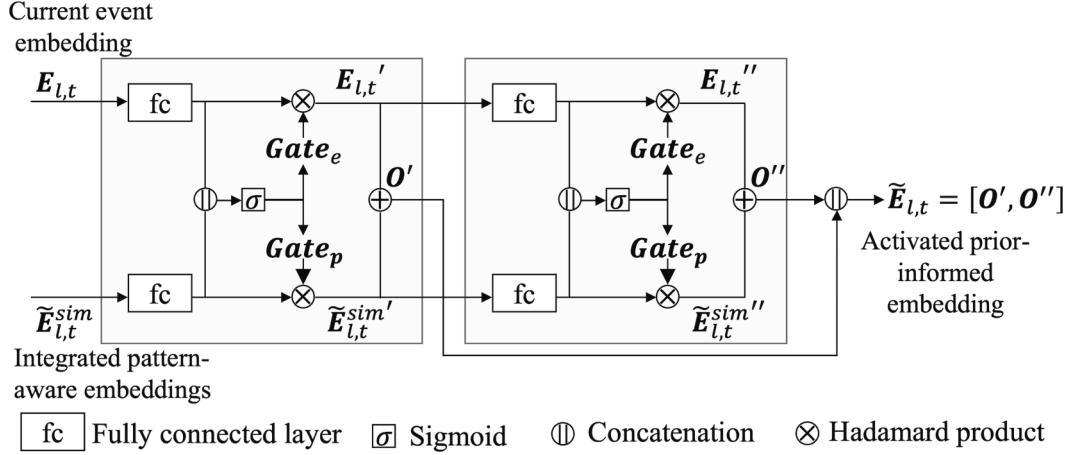


Fig. 10. Prior activation layers.

$$\tilde{e}_{l,(x_i,y_i,t_i)}^{sim} = \sum_0^{n_{ref}} p_{i,j} e_{l,(x_j,y_j,t_j)}^{sim} W_g + b_g \quad (9)$$

where  $\tilde{e}_{l,(x_i,y_i,t_i)}^{sim} \in \mathbb{R}^{c_l}$  is the integrated pattern-aware embedding of  $event_l(x_i, y_i, t_i)$ , and  $e_{l,(x_j,y_j,t_j)}^{sim}$  denotes the embedding of the  $j$ th similar event of  $event_l(x_i, y_i, t_i)$ .

The idea of performing the spatio-temporal convolution on the event affinity graph is to perform a data-driven weighted summation of the embeddings of neighboring nodes  $E_{l,t}^{sim}$  based on the attributed spatio-temporal affinity matrix  $P_{l,t}$ . The result of the spatio-temporal convolution  $\tilde{E}_{l,t}^{sim} \in \mathbb{R}^{c_s \times (n_w \times n_h)}$  integrates the comprehensive correlations between similar events, thereby effectively capturing spatio-temporal pattern priors of historical rainstorms.

Fig. 9 illustrates the procedure of computing the integrated pattern-aware embeddings. Based on the attributed spatio-temporal affinity matrix  $P_{l,t}$  and the embeddings of similar historical events  $E_{l,t}^{sim}$ , spatio-temporal graph convolution is performed to aggregate the embeddings of similar historical events. Note that a fully connected layer is used to align the dimension of  $E_{l,t}^{sim}$  with the results of the spatio-temporal graph convolution. The results of the fully connected layer and the spatio-temporal graph convolution are summed to generate more meaningful

pattern-aware embeddings. Finally, leakyRELU can be used as the activation function to obtain the aggregated integrated pattern-aware embeddings  $\tilde{E}_{l,t}^{sim} \in \mathbb{R}^{c_s \times (n_w \times n_h)}$ .

### 3.3.4. Prior activation layers

The predictions of future rainstorm events depend on the embeddings of both current event and historical events. To effectively fuse these embeddings, we develop prior activation layers to activate the two embeddings as two types of attributed spatio-temporal priors. As shown in Fig. 10, there are two activation layers that learn activated embeddings from  $E_{l,t}$  and  $\tilde{E}_{l,t}^{sim}$ . The resulting activated embeddings are then concatenated to obtain the activated prior-informed embeddings  $\tilde{E}_{l,t}$ . The computation of the first prior activation layer is described as follows:

$$\left\{ \begin{array}{l} Gate_e, Gate_p = \sigma \left( \left[ E_{l,t}, \tilde{E}_{l,t}^{sim} \right] W_a + B_a \right) \\ E_{l,t}' = Gate_e^\circ E_{l,t} \\ \tilde{E}_{l,t}^{sim'} = Gate_p^\circ \tilde{E}_{l,t}^{sim} \\ O' = E_{l,t}' + \tilde{E}_{l,t}^{sim'} \end{array} \right. \quad (10)$$

**Table 2**

Variables of the two datasets used in the experiments.

ERA5 dataset		NCEP dataset	
Variable	Unit	Variable	Unit
Total precipitation	m	Total precipitation	kg m <sup>-2</sup>
2 m temperature	K	Temperature	K
2 m dewpoint temperature	K	Dewpoint temperature	K
Mean sea level pressure	Pa	Pressure reduced to mean sea level	Pa
Surface Pressure	Pa	u-component of wind	m s <sup>-1</sup>
10 m u-component of wind	m s <sup>-1</sup>	v-component of wind	m s <sup>-1</sup>
10 m v-component of wind	m s <sup>-1</sup>	Relative humidity	%
100 m u-component of wind	m s <sup>-1</sup>		
100 m v-component of wind	m s <sup>-1</sup>		

where  $\mathbf{W}_a, \mathbf{B}_a$  are learnable parameters.  $\mathbf{Gate}_e, \mathbf{Gate}_p$  are the learnable activation gates for the embeddings of current events and historical events, respectively.  $\sigma$  denotes the sigmoid activation function.  $\circ$  represents the Hadamard product.

The inputs of the prior activation layers include the embeddings of the current events at time  $t$   $E_{l,t}$  ( $l = 1, 2, 3$ ) and their corresponding integrated pattern-aware embeddings  $\tilde{E}_{l,t}^{sim}$ . Firstly, the current event activation gate  $\mathbf{Gate}_e$  and the spatio-temporal prior activation gate  $\mathbf{Gate}_p$  are computed based on the first equation in Eq. (10) ( $\mathbf{Gate}_e, \mathbf{Gate}_p \in \mathbb{R}^{c_a \times (n_w \times n_h)}$ ). Then using the second and third equations in Eq. (10), the two learned activation gates  $\mathbf{Gate}_e$  and  $\mathbf{Gate}_p$  are used to compute the activated event representation  $E_{l,t}'$  and the prior embedding  $\tilde{E}_{l,t}^{sim'}$  by applying the Hadamard product.  $E_{l,t}'$  and  $\tilde{E}_{l,t}^{sim'}$  are used as the inputs for the second layer. Once the computation of the second layer is completed, the activated prior-informed embedding  $\tilde{E}_{l,t} = [\mathbf{O}', \mathbf{O}']$  can be used to predict rainstorm process.

#### 3.4. Loss function

The entire forecasting network is trained end-to-end using three loss functions: a three-level reconstruction loss for SDGiU-EDN to retain detailed rainstorm information, a standard MSE loss to measure prediction errors, and a spatio-temporal coherence loss with the substantial derivative as a soft constraint to generate physically consistent forecasts. The representation and the forecasting modules are jointly trained by minimizing the overall loss.

As the network structure deepens, the representation loses some detailed spatio-temporal and attributed information. The multi-level reconstruction loss is designed to reduce the information loss during the SDGiU-EDN representation. For each layer, the MSE loss is used to align the input features of the encoder ( $[A_{t-m+1}, \dots, A_t], [S_{1,t-m+1}, \dots, S_{1,t}],$  and  $[S_{2,t-m+1}, \dots, S_{2,t}]$ ) with the reconstructed features of the decoder ( $[\hat{A}_{t-m+1}, \dots, \hat{A}_t], [\hat{S}_{1,t-m+1}, \dots, \hat{S}_{1,t}],$  and  $[\hat{S}_{2,t-m+1}, \dots, \hat{S}_{2,t}]$ ), ensuring that the representation results preserve the original spatio-temporal and attributed information as much as possible. The three-level reconstruction loss can be written as:

$$\begin{aligned} L_{rec} &= L_{rec1} + L_{rec2} + L_{rec3} \\ &= w_{r0}MSE([A_{t-m+1}, \dots, A_t], [\hat{A}_{t-m+1}, \dots, \hat{A}_t]) \\ &\quad + w_{r1}MSE([S_{1,t-m+1}, \dots, S_{1,t}], [\hat{S}_{1,t-m+1}, \dots, \hat{S}_{1,t}]) \\ &\quad + w_{r2}MSE([S_{2,t-m+1}, \dots, S_{2,t}], [\hat{S}_{2,t-m+1}, \dots, \hat{S}_{2,t}]) \end{aligned} \quad (11)$$

where  $S_{1,t-m+1}, \dots, S_{1,t}$  and  $S_{2,t-m+1}, \dots, S_{2,t}$  are the intermediate features of layer 1 and 2 of the encoder, respectively.  $\hat{S}_{1,t-m+1}, \dots, \hat{S}_{1,t}$  and  $\hat{S}_{2,t-m+1}, \dots, \hat{S}_{2,t}$  are the reconstructed features of layer 2 and 3 of the

decoder, respectively.  $w_{r0}, w_{r1}, w_{r2}$  are weights for integrating the MSE losses at the three layers. Equal weights  $w_{r0} = w_{r1} = w_{r2} = \frac{1}{3}$  were set because this configuration empirically yields the best results.

The spatio-temporal coherence loss is based on the substantial derivative equation, which helps to generate forecasts that are physically consistent with the actual formation, development and dissipation patterns of rainstorms. The actual spatio-temporal variations of precipitation can be computed as,

$$\mathbf{C}_{t+1} = \mathbf{R}_{t+2} - \mathbf{R}_{t+1} + \mathbf{U}_{t+1}grad_x(\mathbf{R}_{t+1}) + \mathbf{V}_{t+1}grad_y(\mathbf{R}_{t+1}) \quad (12)$$

where  $\mathbf{U}_{t+1}$  and  $\mathbf{V}_{t+1}$  are the actual moving velocity components of the rainstorm in the  $x$  and  $y$  directions, respectively. These two vectors are computed using the optical flow method (Farnebäck, 2003) based on the actual precipitation for all grid cells in the study region at two consecutive time steps  $\mathbf{R}_{t+1}$  and  $\mathbf{R}_{t+2}$ . Similarly, the predicted spatio-temporal variations of precipitation can be computed as,

$$\widehat{\mathbf{C}}_{t+1} = \widehat{\mathbf{R}}_{t+2} - \widehat{\mathbf{R}}_{t+1} + \mathbf{U}_{t+1}grad_x(\widehat{\mathbf{R}}_{t+1}) + \mathbf{V}_{t+1}grad_y(\widehat{\mathbf{R}}_{t+1}) \quad (13)$$

where  $\widehat{\mathbf{R}}_{t+2}$  and  $\widehat{\mathbf{R}}_{t+1}$  are the predicted precipitations for all grid cells in the study region at two consecutive time steps. Note that  $\mathbf{U}_{t+1}$  and  $\mathbf{V}_{t+1}$  are still used in Eq. (13) to reduce the errors introduced by the predicted values because the predicted rainfall is close to the actual rainfall under the constraint of the MSE loss.

The spatio-temporal coherence loss is then defined to ensure that the forecasted rainfall is consistent with the actual rainfall in terms of spatial and temporal variations:

$$L_{stc} = MSE(\mathbf{C}_{t+1:t+h-1}, \widehat{\mathbf{C}}_{t+1:t+h-1}) \quad (14)$$

where  $\mathbf{C}_{t+1:t+h-1}$  and  $\widehat{\mathbf{C}}_{t+1:t+h-1}$  denote the spatio-temporal variations of actual and predicted precipitation during  $[t+1, t+h-1]$ , respectively. This loss ensures that the spatial and temporal variations of the forecasting results should be close to the spatio-temporal variations of the actual precipitation.

Finally, the overall representation-prediction loss is defined as the weighted sum of the three-level reconstruction loss of the SDGiU-EDN  $L_{rec}$ , the standard MSE loss that measures the prediction error  $L_{MSE}$ , and the spatio-temporal coherence loss  $L_{stc}$ :

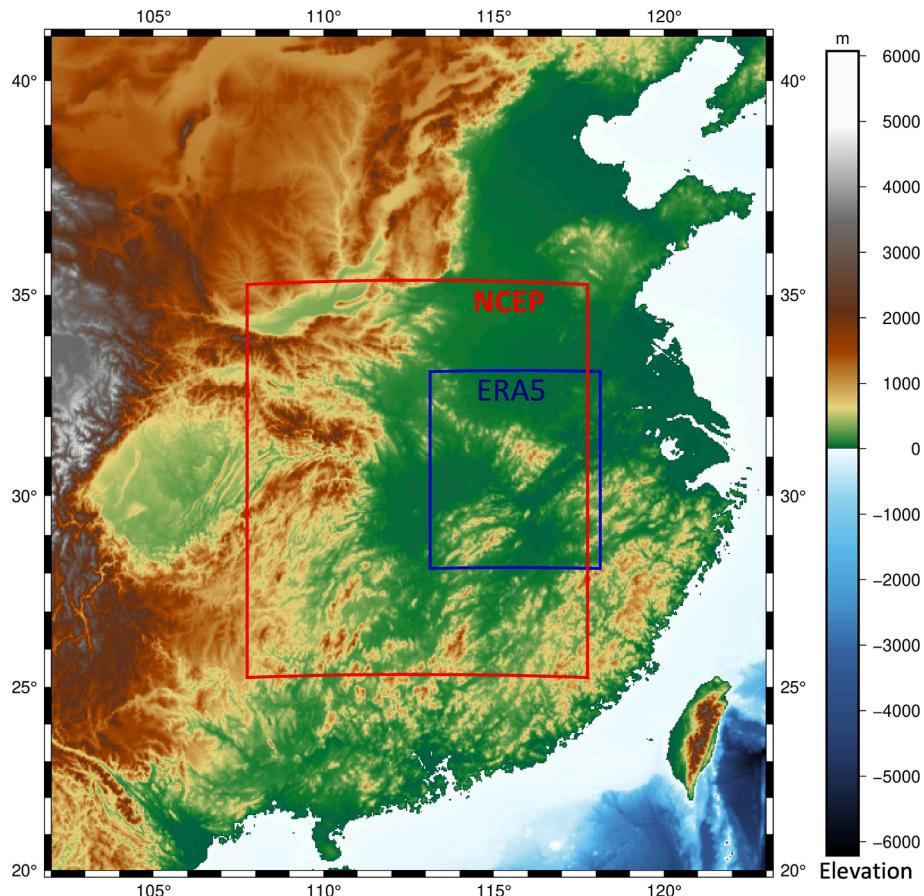
$$\left\{ \begin{array}{l} L_{rec} = w_{r0}MSE([A_{t-m+1}, \dots, A_t], [\hat{A}_{t-m+1}, \dots, \hat{A}_t]) \\ \quad + w_{r1}MSE([S_{1,t-m+1}, \dots, S_{1,t}], [\hat{S}_{1,t-m+1}, \dots, \hat{S}_{1,t}]) \\ \quad + w_{r2}MSE([S_{2,t-m+1}, \dots, S_{2,t}], [\hat{S}_{2,t-m+1}, \dots, \hat{S}_{2,t}]) \\ L_{MSE} = MSE(\mathbf{R}_{t+1:t+h}, \widehat{\mathbf{R}}_{t+1:t+h}) \\ L_{stc} = MSE(\mathbf{C}_{t+1:t+h-1}, \widehat{\mathbf{C}}_{t+1:t+h-1}) \\ L_{com} = w_{ms}L_{MSE} + w_{st}L_{stc} + w_{re}L_{rec} \end{array} \right. \quad (15)$$

where  $w_{ms}, w_{st}, w_{re}$  are weights and  $w_{ms} = w_{st} = 1, w_{re} = 0.1$  were set because this configuration leads to good results in empirical studies.

## 4. Experimental results

### 4.1. Data

Because the quality of observational rainstorm data is not comparable to the reanalysis datasets, which have been widely used in weather forecasting (Scher and Messori, 2019; Rasp et al., 2020; Bi et al., 2023), This study evaluated the proposed rainstorm forecasting model on two widely used reanalysis datasets: the 5th generation European Centre for Medium-range Weather Forecasts Reanalysis dataset (ERA5, Hersbach et al., 2018) and the National Centre for Environmental Prediction Climate Forecast System version 2 selected hourly time-series products (NCEP, Saha et al., 2011). The ERA5 (NCEP) dataset provides data of



**Fig. 11.** Study region. Blue and red boxes represent the spatial extents covered by the ERA5 and NCEP datasets used in this study, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**  
Description of the datasets used in the experiments.

	ERA5	NCEP
Study area	28.125°–33.125° N 113.125°–118.125° E	25.25°–35.25° N 107.75°–117.75° E
Spatial resolution	0.25°	0.5°
Temporal resolution	1 h	1 h
Year of the training data	2010–2016	2011–2017
Year of the validation data	2017	2018
Year of the test data	2019	2019

atmospheric variables on a global scale from 1959 (1979) to the present and has been widely utilized in various studies, including data-driven weather forecasting (Rasp et al., 2020; Pathak et al., 2022; Bi et al., 2023) and climate analysis (Steinhoff et al., 2018; Maillard et al., 2022). The two datasets contain fundamental atmospheric properties such as temperature, pressure, humidity, and wind speed. Table 2 shows the meteorological variables used for event representation.

The spatial extent is presented in Fig. 11. Since most of the rainstorms in the study region occur during the summer months, data from June to August over multiple years was used as the training dataset, one year's data as the validation dataset, and data from another year as the test dataset. The description of the data is given in Table 3.

#### 4.2. Comparison results

The model was implemented in pytorch and trained using an NVIDIA GTX 2080Ti GPU. The initial learning rate was set to  $1 \times 10^{-4}$ . The performance of rainfall prediction was evaluated using Mean Absolute

Error (MAE).

$$MAE = \frac{1}{N_n} \sum_n \frac{1}{N \times h} \sum_{t,i} |r_{t,i,n} - \hat{r}_{t,i,n}| \quad (16)$$

where  $r_{t,i,n}$ ,  $\hat{r}_{t,i,n}$  denote actual and predicted rainfall amounts of the  $i$ th grid at time  $t$  of the  $n$ th test sample,  $N$  is the number of grids in the test dataset,  $N_n$  is the total number of samples in the test dataset, and  $h$  is the number of prediction horizons.

The MAE metric is not sufficient to evaluate the performance of rainstorm prediction due to the low occurrence rate of rainstorms. Therefore, two other metrics, i.e., Critical Success Index (CSI) (Gilbert 1884) and Heidke Skill Score (HSS) (Heidke, 1926), were also used to evaluate the performance of the proposed forecasting model. Higher values of CSI and HSS indicate better performance.

$$CSI = \frac{TP}{TP + FP + FN} \quad (17)$$

$$HSS = \frac{2 \times (TP \times TN - FN \times FP)}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)}$$

where TP is the number of correctly predicted rainstorms. FP is the number of non-rainstorm events that were incorrectly predicted as rainstorms. FN is the number of rainstorm events that were not predicted by the model. TN denotes the number of non-rainstorm events that were correctly predicted.

We compare the forecasting performance of the proposed method with the following baseline models:

- (1) Simmim-Swin (Xie et al., 2022). This baseline is a state-of-the-art masked image modeling framework based on the Swin

**Table 4**

Performance comparison of PGCN and six baseline models on the ERA5 dataset. In addition to MAE between the predicted precipitation and the ground truth, we also evaluated them with the total and step-by-step CSI and HSS metrics.

	*MAE ↓	Total CSI ↑	Total HSS ↑	CSI ↑			HSS ↑		
				<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3
ERA5(16.0 mm/h)									
CNN	0.843	0.264	0.413	0.497	0.218	0.063	0.661	0.353	0.114
TrajGRU	0.922	0.162	0.274	0.313	0.114	0.030	0.472	0.201	0.057
FourcastNet	<b>0.840</b>	0.182	0.304	0.376	0.129	0.011	0.543	0.225	0.022
STSGCN	0.939	0.191	0.314	0.274	0.215	0.047	0.424	0.347	0.086
Simmim-Swin	0.983	0.227	0.366	0.435	0.149	0.039	0.603	0.255	0.073
TCLR	1.329	0.126	0.221	0.270	0.040	0.000	0.420	0.075	0.000
PGCN	0.864	<b>0.316</b>	<b>0.475</b>	<b>0.534</b>	<b>0.273</b>	<b>0.119</b>	<b>0.693</b>	<b>0.423</b>	<b>0.207</b>
ERA5(20.5 mm/h)									
CNN	0.843	0.178	0.300	0.421	0.077	0.000	0.591	0.142	0.000
TrajGRU	0.922	0.079	0.146	0.195	0.000	0.000	0.324	0.000	0.000
FourcastNet	<b>0.840</b>	0.103	0.187	0.273	0.026	0.000	0.427	0.050	0.000
STSGCN	0.939	0.043	0.079	0.048	0.067	0.000	0.089	0.122	0.000
Simmim-Swin	0.983	0.138	0.242	0.352	0.024	0.000	0.519	0.045	0.000
TCLR	1.329	0.074	0.136	0.190	0.000	0.000	0.318	0.000	0.000
Ours	0.864	<b>0.235</b>	<b>0.379</b>	<b>0.494</b>	<b>0.176</b>	<b>0.021</b>	<b>0.660</b>	<b>0.297</b>	<b>0.040</b>

\* Because the MAE metric is computed using absolute precipitation values, it remains consistent across different rainstorm thresholds.

**Table 5**

Performance comparison PGCN and six baseline models on the NCEP dataset.

	MAE ↓	Total CSI ↑	Total HSS ↑	CSI ↑			HSS ↑		
				<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3
NCEP(16.0 mm/h)									
CNN	<b>0.533</b>	0.103	0.186	0.273	0.000	0.000	0.427	0.000	0.000
TrajGRU	1.034	0.039	0.073	0.091	0.000	0.000	0.164	0.000	0.000
FourcastNet	0.648	0.020	0.038	0.053	0.000	0.000	0.099	0.000	0.000
STSGCN	0.971	0.030	0.054	0.024	0.022	0.041	0.044	0.040	0.075
Simmim-Swin	0.634	0.100	0.180	0.189	0.042	0.000	0.316	0.078	0.000
TCLR*	1.424	—	—	—	—	—	—	—	—
Ours	0.713	<b>0.218</b>	<b>0.356</b>	<b>0.429</b>	<b>0.129</b>	<b>0.107</b>	<b>0.599</b>	<b>0.226</b>	<b>0.191</b>

\* Because TCLR failed to predict any rainstorms, it did not report any CSI and HSS metrics. Since there was only one heavy rainstorm ( $\geq 20.5\text{mm/h}$ ) in the NCEP test dataset and none of the baseline methods were able to predict it, the prediction results for the heavy rainstorm are not included in Table 5.

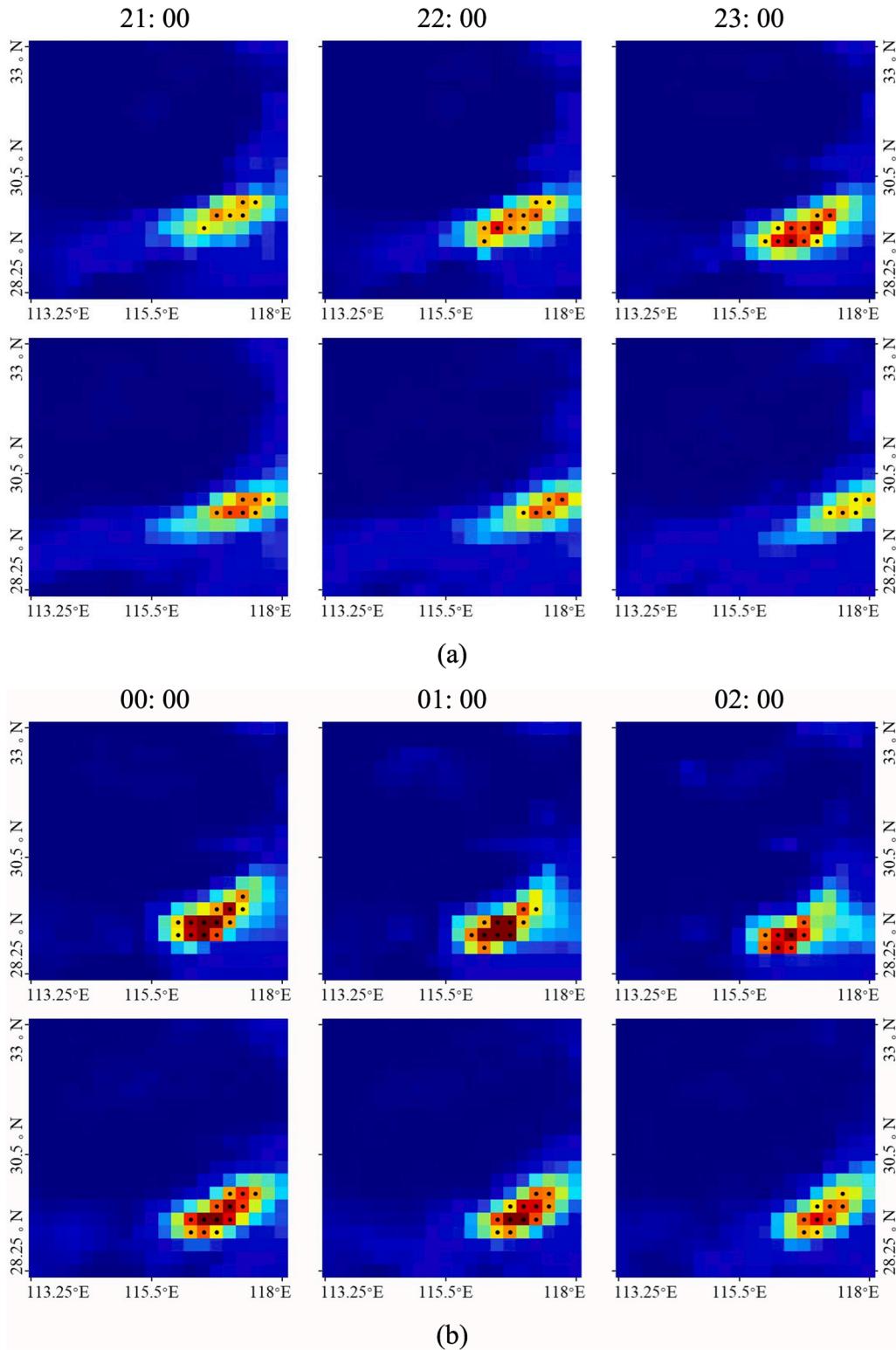
Transformer backbone (Liu et al., 2021). It performs well in the classification task on ImageNet-1 K;

- (2) TCLR(Dave et al., 2022). TCLR is a contrastive learning-driven self-supervised video representation model that uses specially-designed local-local and global-local temporal contrastive losses to promote representation capability. TCLR achieves competitive results in several video understanding tasks;
- (3) CNN (Zhou et al., 2019). CNN is a six-layer convolutional neural network for forecasting severe convective weather based on numerical weather prediction data. It has been practically used as an operational model in the National Meteorological Center of China;
- (4) TrajGRU (Shi et al., 2017). TrajGRU is a classical precipitation nowcasting model with an encoder-forecasting architecture that captures local spatio-temporal correlations through structured recurrent connections;
- (5) FourCastNet (Pathak et al., 2022). FourCastNet is a state-of-the-art global high-resolution weather forecasting model that is able to predict extreme weather events. It is claimed to achieve comparable performance to state-of-the-art numerical weather prediction models;
- (6) STGCN (Song et al., 2020). STGCN is a typical spatio-temporal prediction model that uses spatial-temporal synchronous graph convolutional networks to capture local spatio-temporal correlations and achieves state-of-the-art prediction performance among existing graph neural networks.

The first two baselines are recent advanced visual representation networks. The goal of the performance comparison with these two baselines is to evaluate the effectiveness of the proposed prior-guided representation network SDGiU-EDN. The remaining baselines are typical recent spatio-temporal predictors, some of which have been used for weather forecasting.

Tables 4 and 5 present the rainstorm prediction results for the next 3 hours on the ERA5 and NCEP datasets. Note that “Total CSI” and “Total HSS” were computed over the entire 3-step prediction horizon. Rainstorm and heavy rainstorm events were identified using thresholds of 16 mm/h and 20.5 mm/h, respectively. The results on the ERA5 dataset is presented in Table 4, which indicate that the representation capability of SDGiU-EDN helps the proposed model to achieve the best MAE as well as the total and step-by-step CSI and HSS metrics. For example, the Total CSI is improved by 39.2 % and 150.8 % compared to Simmim-Swin and TCLR, respectively. This performance improvement can be attributed to representation capability of SDGiU-EDN, which effectively integrates atmospheric properties and spatio-temporal statistics of the rainstorm events into compact embeddings.

According to Table 4, the proposed model achieves the best CSI and HSS performance. Compared to CNN, TrajGRU, FourcastNet, and STGCN. It improves the overall CSI improvements by 19.7 %, 95.1 %, 73.6 %, and 65.4 %, respectively. Compared to CNN, the stepwise CSI of the prosed model is improved by 7.4 %, 25.2 %, and 88.9 % on steps 1, 2, and 3, respectively. CNN has the best overall performance among all the compared baseline methods. CNN and FourcastNet have relatively small



**Fig. 12.** Comparison of actual (upper row) and predicted (bottom row) rainfalls over the formation (a), development (b), dissipation (c), and termination (d) of a rainstorm process from 21:00 July 5, to 7:00 July 6, 2018. Grid cells with a black dot indicate rainstorm events (precipitation exceeds the threshold of 16 mm/h).

precipitation errors (MAE), but their rainstorm prediction results are not as good, suggesting a tendency to underestimate extreme rainstorm events.

In Table 4, it can be seen that when the precipitation threshold is 20.5 mm/h, all methods perform worse than when the threshold is 16

mm/h. This indicates that predicting heavy rainstorms becomes more challenging when they are rarer. In this case, the proposed method still provides the best CSI and HSS performance. The total CSI is improved by 70.3 % and 217.6 % compared to Simmim-Swin and TCLR, respectively. While all the baseline methods fail to predict the heavy rainstorms three

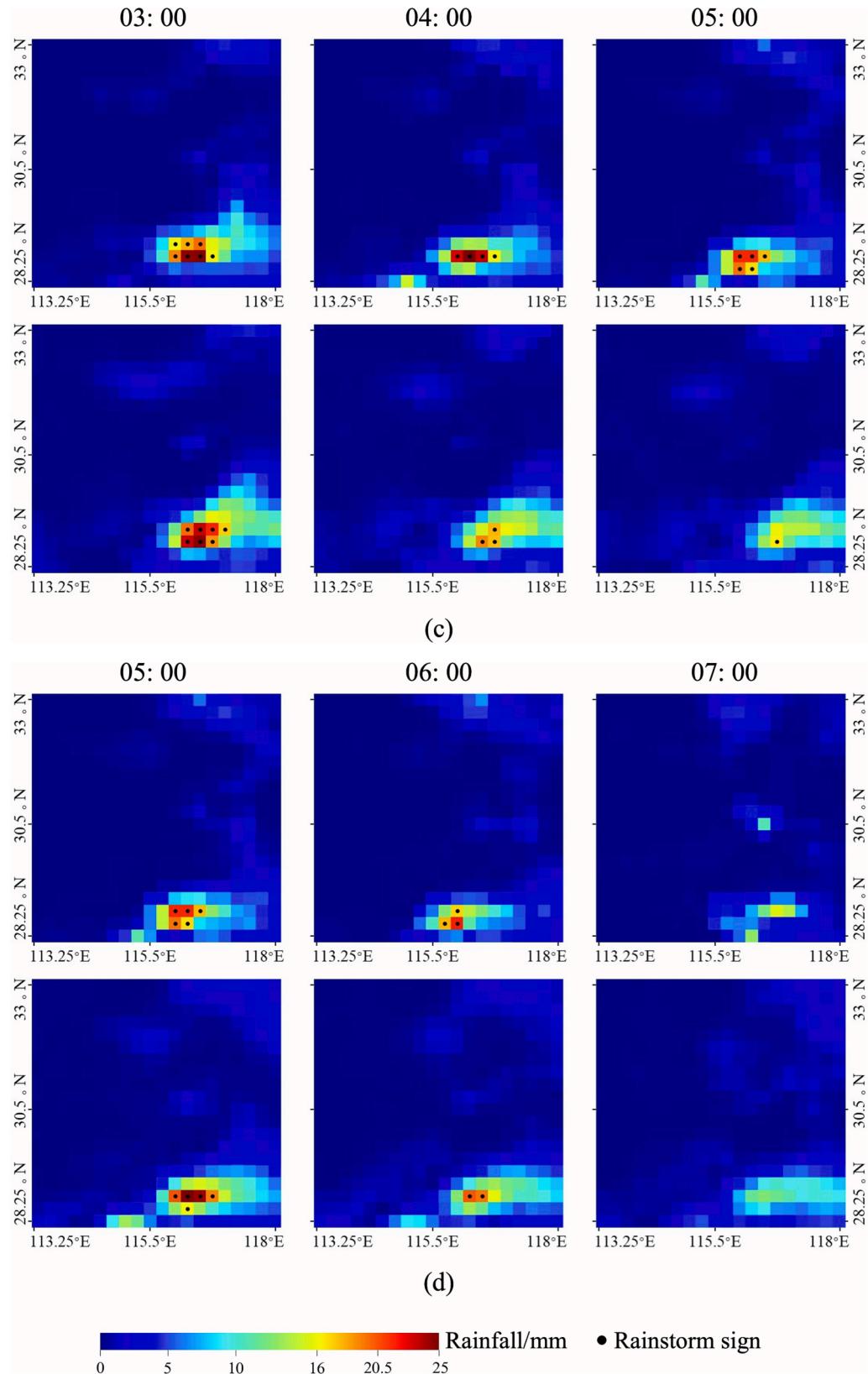


Fig. 12. (continued).

steps ahead ( $h = 3$ ), the proposed method manages to make some correct predictions, suggesting that our method performs better in capturing spatio-temporal correlation of meteorological events with very limited rainstorm samples.

According to the prediction results reported in Table 5, it can be noted that the overall performance of all methods on the NCEP dataset is worse than that of ERA5, probably due to the relatively smaller rainstorm samples. The proposed forecasting method outperforms all the

**Table 6**

Performance comparison of the full PGCN model and four variant models on the ERA5 dataset at two rainstorm thresholds (16 mm/h and 20.5 mm/h) in terms of overall prediction accuracy and step-by-step accuracy.

	MAE ↓	Total CSI ↑	Total HSS ↑	CSI ↑			HSS ↑		
				<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3
ERA5(16.0 mm/h)									
-R-SDGiU	<b>0.764</b>	0.307	0.466	<b>0.592</b>	0.234	0.069	<b>0.742</b>	0.376	0.126
-SDGiU	0.824	0.289	0.443	0.533	0.240	0.118	0.692	0.381	0.205
-STCLoss	0.816	0.272	0.424	0.485	0.235	0.075	0.650	0.375	0.135
-APIE	0.846	0.290	0.446	0.490	0.246	0.109	0.654	0.389	0.191
Full model	0.864	<b>0.316</b>	<b>0.475</b>	0.534	<b>0.273</b>	<b>0.119</b>	0.693	<b>0.423</b>	<b>0.207</b>
ERA5(20.5 mm/h)									
-R-SDGiU	<b>0.764</b>	0.218	0.356	<b>0.529</b>	0.098	0.000	<b>0.690</b>	0.177	0.000
-SDGiU	0.824	0.221	0.360	0.476	0.161	<b>0.036</b>	0.643	0.275	<b>0.067</b>
-STCLoss	0.816	0.187	0.314	0.418	0.104	0.000	0.588	0.187	0.000
-APIE	0.846	0.220	0.359	0.461	0.168	0.007	0.629	0.285	0.012
Full model	0.864	<b>0.235</b>	<b>0.379</b>	0.494	<b>0.176</b>	0.021	0.660	<b>0.297</b>	0.040

**Table 7**

Performance comparison of the full PGCN model and four variant models on the NCEP dataset at the 16 mm/h rainstorm thresholds in terms of overall prediction accuracy and step-by-step accuracy.

	MAE ↓	Total CSI ↑	Total HSS ↑	CSI ↑			HSS ↑		
				<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 3
NCEP(16.0 mm/h)									
-R-SDGiU	0.715	0.183	0.307	0.283	<b>0.135</b>	<b>0.108</b>	0.438	<b>0.235</b>	<b>0.192</b>
-SDGiU	<b>0.625</b>	0.155	0.267	0.346	0.000	0.000	0.513	0.000	0.000
-STCLoss	0.660	0.104	0.187	0.166	0.115	0.000	0.284	0.205	0.000
-APIE	0.648	0.194	0.334	0.389	0.000	0.000	0.558	0.000	0.000
Full model	0.713	<b>0.218</b>	<b>0.356</b>	<b>0.429</b>	0.129	0.107	<b>0.599</b>	0.226	0.191

\* Since there was only one heavy rainstorm ( $\geq 20.5\text{mm/h}$ ) in the NCEP test dataset and none of the variants were able to predict it, the prediction results for the heavy rainstorm are not included in Table 7.

baselines. Compared to the total CSI in Table 4, the prediction performance based on Simmim-Swin is reduced by 55.9 % in the NCEP dataset. TCLR is unable to predict any rainstorm. The performance of the SDGiU-EDN-based predictor is reduced by only 31.0 %, suggesting that the scarcity of training samples has less negative effect on the proposed method. Although Simmim-Swin achieved the best MAE, it performed much worse in terms of CSI and HSS, which are more informative metrics for rainstorm prediction than MAE.

The comparison experiments demonstrate that the proposed method outperforms all baselines on both the ERA5 and NCEP datasets at two rainstorm thresholds (16 mm/h and 20.5 mm/h), in terms of overall prediction accuracy as well as step-by-step accuracy. Two key mechanisms of the proposed method may contribute to the performance improvement: (1) the effective integration of substantial derivatives of meteorological variables in the gated convolutional network; and (2) the full integration of spatio-temporal and attributed information via the multi-layer reconstruction loss and the spatio-temporal coherence loss. The two developments empower the predictor with the ability to characterize the spatio-temporal dynamics of rainstorms with limited training data. More experimental results can be found in the Appendices.

#### 4.3. Qualitative analysis

Fig. 12 illustrates an example of how the proposed model can predict the process of formation, development, dissipation and termination of a rainstorm, which occurred during the 2018 monsoon season. It can be seen that the coverage of rainstorm increased/decreased rapidly as the rainstorm developed/dissipated, making it challenging to capture changes in rainstorm events. From this example, it can be seen that the proposed model successfully predicts the entire lifecycle of a rainstorm event including formation, development, dissipation, and

termination. It can accurately predict the coverage of rainstorms up to 1 h.

In July, rainstorms are likely to occur under sufficient water vapor conditions due to the convergence of the south-west monsoon and cold air from the north (Guan et al., 2020). After 20:00 on 5 July, water vapor convergence can be observed in the south-eastern part of the study region, with a low-pressure center in the vicinity of the water vapor convergence zone, which provided favorable water vapor conditions for the formation and development of the rainstorm. The predicted rainstorm area was close to the water vapor convergence zone. After 2:00 on 6 July, the water vapor convergence gradually weakened and was unable to provide sufficient water vapor for the continuous development of the rainstorm. This dissipation of the rainstorm was also successfully predicted by the proposed model. The results indicate that the proposed model can utilize spatio-temporal pattern and substantial derivative priors to guide the gated convolutional network to learn integrated spatio-temporal-attributed representations of rainstorm events even when there are limited training samples. With the substantial derivative equation and the spatio-temporal coherence loss, the model can generate non-linear latent representations to encode complex and meaningful spatio-temporal-attributed correlations.

#### 4.4. Ablation studies

To evaluate the effectiveness of the proposed SDGiU, the spatio-temporal coherence loss, and activated prior-informed embeddings, we performed ablation studies using four variants of the proposed model:

- 1) -R-SDGiU: this variant replaces all SDGiUs in SDGiU-EDN with the classical ConvGRU (Shi et al., 2017);

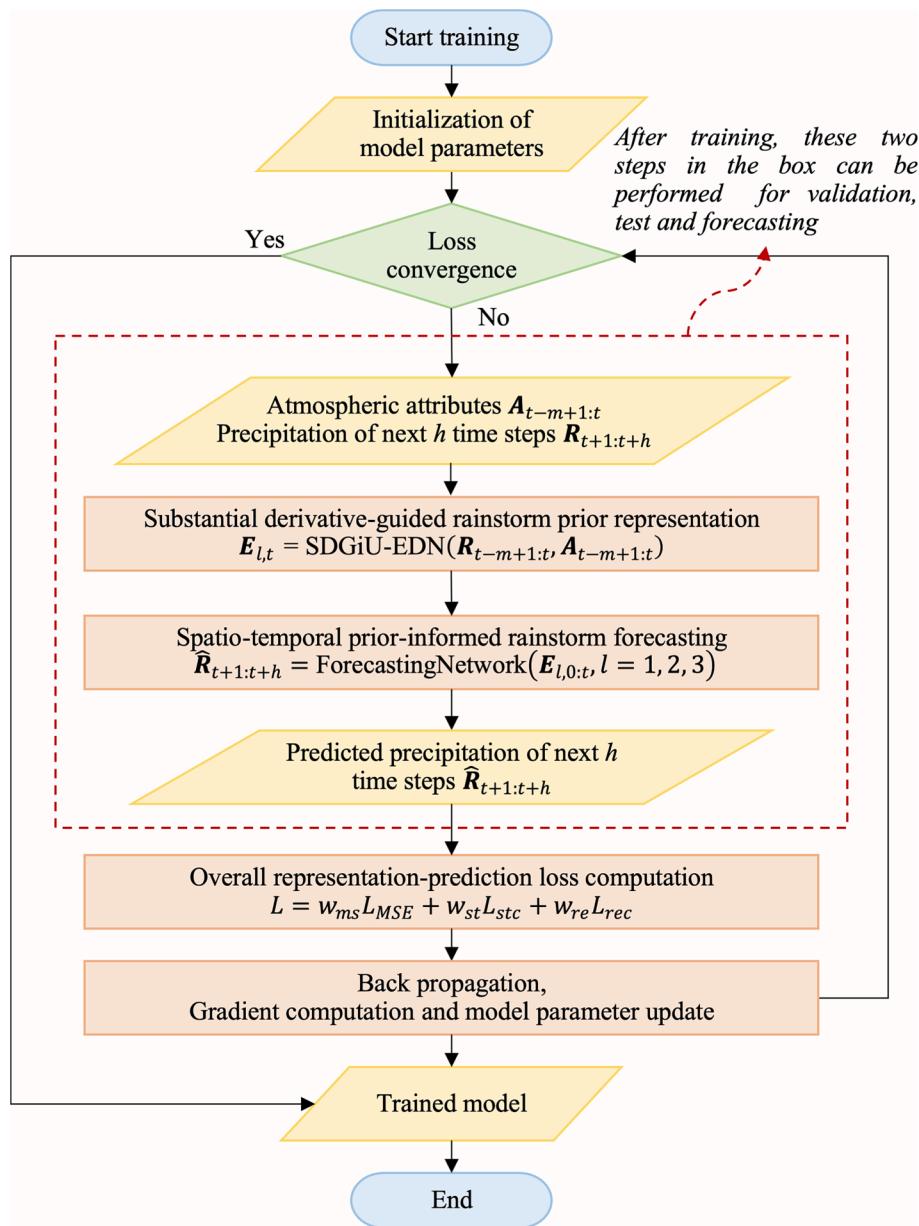


Fig. A1. Workflow of the proposed rainstorm prediction model.

- 2) -SDGiU: this variant replaces all SDGiUs in both SDGiU-EDN and the forecasting network with the classical ConvGRU (Shi et al., 2017);
- 3) -STCLoss: in this variant, the spatial temporal coherence loss is not used.
- 4) -APIE: this variant does not implement spatio-temporal convolution and the prior activation layers, which means Activated Prior-Informed Embeddings are not produced and used in forecasting.

Tables 6 and 7 reports the results of applying the full model and its variants to predict rainstorms for the next 3 hours. Compared to the results of Tables 4 and 5, all four variants of the proposed model show better total CSI and HSS scores than all the compared baselines, proving the validity of the proposed prior-informed modules and the loss function. -STCLoss exhibits relatively large performance degradation, implying that the spatio-temporal coherence loss function plays a more important role than the other components in capturing the spatio-temporal variation patterns of rainstorms.

Table 7 shows that all variants perform worse than the full model, due to very limited sample size of the NCEP dataset. -SDGiU and -APIE

performed worse than other variants in steps 2 and 3, meaning that effective representation and integration of spatio-temporal-attributed rainstorm features are critical for longer horizon predictions. -SDGiU performed worse than -SDGiU at step 1, but better at steps 2 and 3, indicating an overestimation of precipitation. Compared to the results of Table 6, it can be observed that the performance gap between -SDGiU and -APIE is much larger, suggesting that that SDGiUs can use physical priors to mitigate the shortage of rainstorm samples.

The ablation studies show a significant performance degradation when SDGiU is not incorporated in the model, suggesting that the prior representation network, and SDGiU in particular, is vital to the overall performance of the PGCN model, which supports the claim of the second contribution. The integrated loss function also plays a critical role in promoting the forecasting performance (as shown in Tables 6-7) and reproducing the development patterns of rainstorm (as shown in the qualitative analysis in Section 4.3), which corroborates the claim of the third contribution. In summary, the PGCN model can effectively models the substantial derivatives and spatio-temporal statistics of atmospheric dynamics as physical priors. By combining these priors with gate

convolutional networks, it improves sample efficiency and generalizability, which substantiate the claim of the first contribution.

## 5. Discussion and conclusion

We have presented a prior-guided rainstorm forecasting method that integrates learnable event representations based on a gated convolutional network, addressing the challenges of complex spatio-temporal correlations and the heterogeneous influences of atmospheric factors on rainstorm development. The substantial derivative equation is used to guide the embedding of rainstorm events by enforcing physical constraints in specially-designed gated convolutional units, which also incorporate spatio-temporal statistics to capture the spatio-temporal development patterns of historical rainstorms. The experimental results have demonstrated that the proposed prediction model can generate physically more consistent and spatially-temporally more coherent predictions than compared baselines. Although the proposed model was evaluated using the rainstorm data of China, it can be applied worldwide by applying a local criterion to identify rainstorm events and using them as training samples. The proposed prior-guided gated convolutional network is built based on generic physical priors and widely used deep learning models, making it generalizable to data collected from other countries.

The results of this study show that prior-guided machine learning models have the potential to effectively integrate various physical priors into automated data-driven learning frameworks to emulate complex atmospheric dynamics of rainstorms and address the challenges of rainstorm forecasting. It can be expected that more resources, including benchmark datasets, computational systems and domain expertise, will be devoted to develop more robust and scalable prior-guided deep learning models for high-resolution rainstorm modeling and skillful prediction. We also anticipate the prediction accuracy of these models will surpass that of the current NWP models and will be used operationally in the near future. Joint efforts of different disciplines, such as meteorology, computer science, and remote sensing, can be coordinated to achieve this goal. To cost-effectively utilize the accumulated expertise and knowledge of meteorologists, we recommend integrating legacy NWP models with deep learning models.

The main limitations of the proposed methodology are: 1) the case studies were based on reanalysis datasets rather than observational rainstorm data or other meteorological data sources (e.g., radar echo data). There is a risk that realistic rainstorms development patterns may not be fully captured; 2) this study only accounts for some elementary forms of rainstorm priors. The priors of rainstorms are not sufficiently modelled, which may have a negative impact on the generalization ability of the proposed model.

Additional data, such as radar, observational precipitation, and topographical data can be incorporated into the proposed forecasting framework to improve the forecasting performance of rainstorm events. More advanced priors can be modeled and incorporated into data-driven

forecasting frameworks, such as the governing equations of atmospheric motion or statistical constraints based on the spatio-temporal structures of rainstorm processes. Future studies can also explore to combine operational NWP models with the proposed model and evaluate the applicability of such hybrid models for high-resolution rainstorm prediction. In the future, the proposed “event representation-prediction” framework can also be applied to the prediction of other extreme climate events, such as floods or droughts, which is vital to mitigate the impacts of climate change and achieve global Sustainable Development Goals (SDGs) ([United Nations, 2015](#)).

## Funding

This work was supported by the Joint Research Program for Enhancing Meteorological Capabilities by the China Meteorological Administration (grant number 22NLTSY015), Hubei Provincial Natural Science Foundation of China (grant number 2022CFD012), Key R&D Program of Hubei Province (grant number 2023BCB119), China National Postdoctoral Support Program for Innovative Talents (grant number BX20230360), the Open Project Fund of China Meteorological Administration Basin Heavy Rainfall Key Laboratory (grant number 2023BHR-Y13) and Open Fund of State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University (grant number 23I03).

## CRediT authorship contribution statement

**Tong Zhang:** Writing – original draft, Resources, Funding acquisition, Conceptualization. **Jie Liu:** Visualization, Validation, Software, Methodology. **Chulin Gao:** Methodology. **Peixiao Wang:** Writing – review & editing. **Liang Leng:** Writing – review & editing. **Yanjiao Xiao:** Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data Availability Statement The data that support the findings of this study are openly available at <http://doi.org/10.24381/cds.adbb2d47> and <https://doi.org/10.5065/D6N877VB>.

## Acknowledgements

The numerical calculations in the paper have been done on the supercomputing system in the Supercomputing center of Wuhan University.

## Appendix

### 1. Workflow of the entire prediction method

[Fig. A1](#) illustrates the training, validation, and test phrases in an integrated workflow. The training of the model focuses on the two modules of *substantial derivative-guide rainstorm prior representation* and *spatio-temporal prior-informed rainstorm forecasting*. The model is trained by the overall representation-prediction loss. Once the loss converges, the training is stopped and the trained model (outlined with dash line) can be used for validation and testing. During the validation stage, the goal is to search for the best hyper-parameters using a validation dataset. Once the configuration of hyper-parameters is determined, we can proceed to test the model and produce forecasting for the next  $h$  hours based on the inputs of atmospheric attributes of previous  $m$  hours. In summary, the trained representation and forecasting modules are used in the validation and testing phrases. The computation of loss and back-propagation is performed only in the training stage.

[Fig. A1](#)

## 2. Substantial derivatives as physical priors

In fluid modeling, the substantial derivative is used to describe the rate of change of a physical quantity along a particular path of flowing fluid (Anderson, 1992). As shown in Eq. (A1), the substantial derivative of the physical quantity  $H$  in 3D Cartesian coordinates ( $x, y, z$ ) can be written as a function of  $x, y, z$ , and  $t$ ,

$$\frac{dH}{dt} = u \frac{\partial H}{\partial x} + v \frac{\partial H}{\partial y} + w \frac{\partial H}{\partial z} + \frac{\partial H}{\partial t} \quad (\text{A1})$$

where  $u, v$  and  $w$  represent the velocity components in  $x, y$ , and  $z$  directions, respectively.  $\frac{\partial H}{\partial t}$  denotes the local derivative, which represents the change rate observed from a fixed location in the 3D coordinate system.  $u \frac{\partial H}{\partial x} + v \frac{\partial H}{\partial y} + w \frac{\partial H}{\partial z}$  can be regarded as the convective derivative, which describes the change caused by the transport of the physical quantity in the fluid. Eq. (A1) can be also simplified as,

$$\frac{dH}{dt} = (\mathbf{V} \bullet \nabla)H + \frac{\partial H}{\partial t}, \quad \left( \nabla H = i \frac{\partial H}{\partial x} + j \frac{\partial H}{\partial y} + k \frac{\partial H}{\partial z} \right) \quad (\text{A2})$$

where  $\mathbf{V}$  is a velocity vector with the components of  $u, v$  and  $w$  in  $x, y$ , and  $z$  directions, respectively.  $\nabla$  denotes the gradient operator.

From Eqs. (A1) and (A2), it can be seen that the substantial derivative of a physical quantity can be computed as the summation of local and convective derivatives. Some recent data-driven weather forecasting studies have incorporated the basic principle of the substantial derivative equation into deep learning models to better capture the spatio-temporal variability of meteorological quantities. For example, Bézenac et al. (2019) develop a differentiable warping scheme based on the advection-diffusion equation for predicting future sea surface temperature.

The concept and equations of the substantial derivative can be used to describe the fundamental physical laws governing the motion of atmospheric fluid. Since many meteorological quantities vary in time and space, the substantial derivative equations are used to represent the fundamental physical prior, which describes how atmospheric dynamics change along the path of a rainstorm. This study incorporates them into a data-driven deep rainstorm forecasting model, with the aim of improving the capability to capture dynamic motion patterns and spatio-temporal correlations of rainstorms.

The integration of substantive derivatives also contributes to the generation of physically consistent forecasts by constraining the model to obey basic physical principles that govern atmospheric dynamics. Specifically, a substantial derivative-guided gated convolutional encoder-decoder network is developed to effectively represent historical rainstorm events and alleviate the problem of limited training samples. This study also develops a spatio-temporal coherence loss with the substantial derivative as a soft constraint to generate forecasts that are more consistent with the spatial and temporal patterns of the rainstorm event development process.

## 3. Computation of SDGiU-EDN

Using the atmospheric attributes (including precipitation data) of the last  $m$  steps as input, SDGiU-EDN produces embeddings of historical rainstorm events from the training dataset. Once the training is completed, the encoder of SDGiU-EDN (Lines 1–5 in Algorithm A1) can be used to generate representations for both non-rainstorm and rainstorm events for the last  $m$  steps before the current time  $t$ . The decoder (Lines 6–10 in Algorithm A1) reconstructs the vectors of atmospheric attributes.

---

Algorithm A1. Substantial derivative-guided event representation (SDGiU-EDN)

---

```

Input:  $A_{t-m+1}, \dots, A_t$ 
      //atmospheric attribute data of the last  $m$  steps (including precipitation data)
       $F_t$  //spatio-temporal statistics of rainstorms of the last  $m$  steps
Output:
       $E_{lt}, l = 1, 2, 3$  // learned event embedding
       $\hat{A}_{t-m+1}, \dots, \hat{A}_t$  // reconstructed atmospheric attribute data(including precipitation data)
      1  $A = A_{t-m+1}$ 
      2 for  $l=1$  to 3 step 1 // the encoder has three layers
      3  $A, H := \text{SDGiU}[l](A, 0)$  //hidden state matrix is initialized to all zeros
      4  $E_{lt} := \text{conv}(H + F_t)$  //compute event embeddings
      5 end for
      6  $A := \text{None}$ 
      7 for  $l=3$  to 1 step -1 // the decoder has three layers
      8  $A, H := \text{SDGiU}[l](A, E_{lt})$ 
      9 end for
      10  $\hat{A}_{t-m+1}, \dots, \hat{A}_t := A$ 

```

---

Each layer of the encoder and decoder consists of  $m$  recurrent SDGiUs, which are guided by the substantial derivative equation to update the temporal hidden states. The specific procedure of one SDGiU module in an encoder/decoder layer is illustrated in Algorithm A2. Each layer consists of  $m$  recurrent SDGiUs (line 3). In line 4, the algorithm computes the residual matrix  $Res$  and the velocity components in the hidden state in  $x$  and  $y$  directions ( $U, V$ ). In line 5, the state is updated based on  $Res$  and  $(U, V)$  according to Eq. (5). Lines 6–9 implement Eq. (6) to compute the update and reset gates and produce the final hidden state representation of the SDGiU module.

---

Algorithm A2. Computation of the SDGiU ( Substantial derivative-guided gated convolutional unit ) module within each encoder/decoder layers

---

```

Input:  $A_{t-m+1}, \dots, A_t$   

//atmospheric attribute data of the last  $m$  steps (including precipitation data)

Output:
 $H_1, \dots, H_m$  // Hidden states of each SDGiU,  $H_m$  is the final hidden state of the current layer
1  $A_{1:m} := A_{t-m+1:t}$ 
2  $H_0 := 0$  // hidden state matrix is initialized to all zeros
3 for  $l=1$  to  $m$  //each module has  $m$  recurrent SDGiUs
4  $U, V, Res := uvr(A_l, H_{l-1})$ 
5  $H'_{l-1} := H_{l-1} - U^T \text{grad}_x(H_{l-1}) - V^T \text{grad}_y(H_{l-1}) - Res$ 
6  $Z_l := \sigma(\text{conv}_{az}(A_l) + \text{conv}_{hz}(H'_{l-1}))$ 
7  $R_l := \sigma(\text{conv}(A_l) + \text{conv}_{hr}(H'_{l-1}))$ 
8  $\tilde{H}_l := t(\text{conv}_{ah}(A_l) + R_l^T \text{conv}_{hh}(H'_{l-1}))$ 
9  $H_l := (1 - Z_l)^\circ \tilde{H}_l + Z_l H_{l-1}$ 
10 end for
11  $A_{1:m} := H_{1:m}$  // the final hidden state of each SDGiU module is fed into the next layer as input

```

---

#### 4. More experimental results

To comprehensively evaluate the performance of the proposed model, this study compares it to the baselines in terms of Probability of Detection (POD), Miss Alarm Rate (MAR) and False Alarm Rate (FAR) (Eq. A(3)). POD is the rate of the number of correctly predicted rainstorms to the number of actual rainstorms. MAR is the rate of the number of rainstorm events that were not predicted by the model to the number of rainstorms. FAR is the rate of the number of non-rainstorm events that were incorrectly predicted as rainstorms to the number of predicted rainstorms. Higher POD values and lower MAR/FAR values indicate better performance.

$$\begin{aligned}
POD &= \frac{TP}{TP + FN} \\
MAR &= \frac{FN}{TP + FN} \\
FAR &= \frac{FP}{TP + FP}
\end{aligned} \tag{A3}$$

where TP is the number of correctly predicted rainstorms. FP is the number of non-rainstorm events that were incorrectly predicted as rainstorms. FN is the number of rainstorm events that were not predicted by the model. TN denotes the number of non-rainstorm events that were correctly predicted. Note that  $POD + MAR = 1$ .

**Table A1**  
Comparison of prediction reliability on ERA5.

	POD $\uparrow$	MAR $\downarrow$	FAR $\downarrow$
<b>ERA5(16 mm/h)</b>			
CNN	0.326	0.674	0.385
TrajGRU	0.201	0.799	0.551
FourcastNet	0.199	0.801	<b>0.327</b>
STGCN	0.299	0.701	0.656
Simmim-Swin	0.288	0.712	0.484
TCLR	0.080	0.920	0.580
PGCN	<b>0.444</b>	<b>0.556</b>	0.478

**Table A2**  
Comparison of prediction reliability on NCEP.

	POD $\uparrow$	MAR $\downarrow$	FAR $\downarrow$
<b>NCEP(16.0 mm/h)</b>			
CNN	0.125	0.875	<b>0.625</b>
TrajGRU	0.063	0.936	0.906
FourcastNet	0.021	0.979	0.750
STGCN	0.083	0.917	0.956
Simmim-Swin	0.167	0.833	0.800
TCLR	-	-	-
PGCN	<b>0.396</b>	<b>0.604</b>	0.672

According to Tables A1 and A2, the proposed PGCN model substantially outperforms the comparison method in terms of POD and MAR, indicating a relatively high detection rate and low miss rate. However, the proposed method does not perform as well as CNN or FourcastNet in the FAR metric, suggesting a tendency to overestimate precipitation compared to CNN and FourcastNet. The proposed method has the best performance in CSI, which is a widely used metric in rainstorm prediction. Therefore, we argue that the proposed method has achieved a better balance in terms of accuracy, false alarm, and miss alarm. In future research, we will continue to improve the rainstorm prediction performance by reducing the false alarm rate.

## References

- China Meteorological Administration. 2012. Definition and classification of rainstorm [online]. Available from : [http://www.cma.gov.cn/2011qxfw/2011qqxkp/2011qkpd/201205/t20120508\\_172024.html](http://www.cma.gov.cn/2011qxfw/2011qqxkp/2011qkpd/201205/t20120508_172024.html) (In Chinese) [Accessed July 2023].
- Anderson, J.D., 1992. Governing equations of fluid dynamics. In: Wendt, J.F. (Ed.), Computational Fluid Dynamics. Springer, Berlin, Heidelberg.
- Bai, C., Sun, F., Zhang, J., Song, Y., Chen, S., 2022. Rainformer: features extraction balanced network for radar-based precipitation nowcasting. *IEEE Geosci. Remote Sens. Lett.* 19, 4023305.
- Bauer, P., Thorpe, A., Brunet, G., 2015. The quite revolution of numerical weather prediction. *Nature* 525, 47–55.
- Bézenac, E., Pajot, A., Gallinari, P., 2019. Deep learning for physical processes: incorporating prior scientific knowledge. *J. Stat. Mech: Theory Exp.* 12, 124009.
- Bi, K., Xie, L., Zhang, H., Chen, X., Gu, X., Tian, Q., 2023. Accurate medium-range global weather forecasting with 3D neural networks. *Nature* 619, 533–538.
- Buehner, M., Jacques, D., 2020. Non-gaussian deterministic assimilation of radar-derived precipitation accumulations. *Mon. Weather Rev.* 148 (2), 2783–2808.
- Dave, I., Gupta, R., Rizve, M., Shah, M., 2022. TCLR: temporal contrastive learning for video representation. *Comput. Vis. Image Underst.* 219, 103406.
- Farnabæk, G. 2003. Two-frame motion estimation based on polynomial expansion. In: *Image Analysis: 13th Scandinavian Conference, SCIA 2003* Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13 (pp. 363–370). Springer Berlin Heidelberg.
- Gilbert, G., 1884. Finley's Tornado Predictions. *American Meteorological Journal* 1, 166–172.
- Guan, P., Chen, G., Zeng, W., Liu, Q., 2020. Corridors of mei-yu-season rainfall over eastern China. *J. Clim.* 33 (7), 2603–2626.
- Heidke, P., 1926. Calculation of the success and goodness of strong wind forecasts in the storm warning service. *Geogr. Ann.* 8, 301–349.
- Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. 2018. ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Doi: 10.24381/cds.adbb2d47.
- Hess, P., Boers, N., 2022. Deep learning for improving numerical weather prediction of heavy rainfall. *J. Adv. Model. Earth Syst.* 14 (3) e2021MS002765.
- Huang, X., Luo, C., Ye, Y., Li, X., Zhang, B., 2022. Location-refining neural network: a new deep learning-based framework for heavy rainfall forecast. *Comput. Geosci.* 166, 105152.
- IPCC (Intergovernmental Panel on Climate Change). 2022. Climate change 2022: mitigation of climate change (summary for policymakers). [https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC\\_AR6\\_WGIII\\_FullReport.pdf](https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf).
- Karniadakis, G., Kevrekidis, I., Lu, L., Perdikaris, P., Wang, S., Yang, L., 2021. Physics-informed machine learning. *nature reviews. Physics* 3, 422–440.
- Kashinath, K., Mustafa, M., Albert, A., et al., 2021. Physics-informed machine learning: case studies for weather and climate modelling. *Phil. Trans. R. Soc. A* 379 (2194), 20200093.
- Kuriqi, A., Ardiclioğlu, M., 2018. Investigation of hydraulic regime at middle part of the Loire River in context of floods and low flow events. *Pollack Periodica* 13 (1), 145–156.
- Li, W., Jiang, R., Wu, H., Xie, J., Zhao, Y., Song, Y., Li, F., 2023. A system dynamics model of urban rainstorm and flood resilience to achieve the sustainable development goals. *Sustain. Cities Soc.* 96, 104631.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030.
- Liu, J., Xu, L., Chen, N., 2022. A spatiotemporal deep learning model ST-LSTM-SA for hourly rainfall forecasting using radar echo images. *J. Hydrol.* 609, 127748.
- Maillard, L., Boucharel, J., Renault, L., 2022. Direct and rectified effects of tropical instability waves on the eastern tropical Pacific mean state in a regional ocean model. *J. Phys. Oceanogr.* 52 (8), 1817–1834.
- Ornes, S., 2018. How does climate change influence extreme weather? impact attribution research seeks answers. *Proc. Natl. Acad. Sci.* 115 (33), 8232–8235.
- Palmer, T., Shutts, G., Hagedorn, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., 2005. Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.* 33 (1), 163–193.
- Pathak, J., Subramanian, S., Harrington, P., et al. 2022. FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. <https://arxiv.org/abs/2202.11214>.
- Piadeh, F., Behzadian, K., Alani, A., 2022. A critical review of real-time modelling of flood forecasting in urban drainage systems. *J. Hydrol.* 607, 127476.
- Rasp, S., Dueben, P., Scher, S., et al., 2020. WeatherBench: a benchmark data set for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems* 12 (11) e2020MS002203.
- Ravuri, S., Lenc, K., Willson, M., et al., 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597 (7878), 672–677.
- Ritvanen, J., Harnist, B., Aldana, M., Mäkinen, T., Pulkkinen, S., 2023. Advection-free convolutional neural network for convective rainfall nowcasting. *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing* 16, 1654–1667.
- Saha, S., et al., 2011. NCEP climate forecast system version 2 (CFSv2) selected hourly time-series products. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. <https://doi.org/10.5065/D6N877VB>.
- Scher, S., Messori, G., 2019. Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. *Geosci. Model Dev.* 12 (7), 2797–2809.
- Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D., Wong, W., Woo, W., 2017. Deep learning for precipitation nowcasting: a benchmark and a new model. In: *Advances in Neural Information Processing Systems*, p. 30.
- Shortridge, J., 2019. Observed trends in daily rainfall variability result in more severe climate change impacts to agriculture. *Clim. Change* 157 (3–4), 429–444.
- Song, C., Lin, Y., Guo, S., and Wan, H. 2020. Spatial-temporal synchronous graph convolutional networks: a new framework for spatial-temporal network data forecasting. *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI) Conference on Artificial Intelligence* 34(1): 914–921.
- Steinhoff, D., Bruijntjes, R., Hacker, J., et al., 2018. Influences of the monsoon trough and arabian heat low on summer rainfall over the United Arab Emirates. *Mon. Weather Rev.* 146 (5), 1383–1403.
- Sun, J., 2005. Convective-scale assimilation of radar data: progress and challenges. *Q. J. R. Meteorolog. Soc.* 131 (613), 3439–3463.
- United Nations, 2015. Transforming our world: the 2030 agenda for sustainable development. Retrieved from. <https://sdgs.un.org/2030agenda>.
- von Rueden, L., Mayer, S., Bechkh, K., et al., 2021. Informed machine learning: a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE Trans. Knowl. Data Eng.* 33 (1), 614–633.
- Wu, M., Wu, Z., Ge, W., Wang, H., Shen, Y., Jiang, M., 2021. Identification of sensitivity indicators of urban rainstorm flood disasters: a case study in China. *J. Hydrol.* 599, 126393.
- Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., Hu, H., 2022. Simmim: a simple framework for masked image modeling. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2022*, 9653–9663.
- Yang, S., Yuan, H., 2023. A customized multi-scale deep learning framework for storm nowcasting. *Geophys. Res. Lett.* 50 e2023GL103979.
- Zhang, W., Li, W., Zhu, L., Ma, Y., Yang, L., Lott, F.C., Li, C., Dong, S., Tett, S., Dong, B., Sun, Y., 2020. Anthropogenic influence on 2018 summer persistent heavy rainfall in central western China. *Bulletin of the American Meteorological Society* 101 (1), S65–S70.
- Zhang, T., Liu, J., Wang, J., 2022. Rainstorm prediction via a deep spatio-temporal attributed affinity network. *Geocarto Int.* 37 (26), 13079–13097.
- Zhou, K., Zheng, Y., Li, B., Dong, W., Zhang, X., 2019. Forecasting different types of convective weather: a deep learning approach. *J. Meteorolog. Res.* 33 (5), 797–809.